SENSORCOMM 2013

The Seventh International Conference on Sensor Technologies and Applications

ISBN: 978-1-61208-296-7

August 25-31, 2013

Barcelona, Spain

SENSORCOMM 2013 Editors

Sergey Yurish, IFSA - Barcelona, Spain

Muhammad Shakeel Virk, Narvik University College, Norway

SENSORCOMM 2013

Foreword

The Seventh International Conference on Sensor Technologies and Applications [SENSORCOMM 2013], held between August 25-31, 2013 in Barcelona, Spain, continued a series of events covering related topics on theory and practice on wired and wireless sensors and sensor networks.

Sensors and sensor networks have become a highly active research area because of their potential of providing diverse services to broad range of applications, not only on science and engineering, but equally importantly on issues related to critical infrastructure protection and security, health care, the environment, energy, food safety, and the potential impact on the quality of all areas of life.

Sensor networks and sensor-based systems support many applications today on the ground. Underwater operations and applications are quite limited by comparison. Most applications refer to remotely controlled submersibles and wide-area data collection systems at a coarse granularity.

Underwater sensor networks have many potential applications such a seismic imaging of undersea oilfields as a representative application. Oceanographic research is also based on the advances in underwater data collection systems.

There are specific technical aspects to realize underwater applications which can not be borrowed from the ground-based sensors net research. Radio is not suitable for underwater systems because of extremely limited propagation. Acoustic telemetry could be used in underwater communication; however off-the-shelf acoustic modems are not recommended for underwater sensor networks with hundreds of nodes because they were designed for long-range and expensive. As the speed of light (radio) is five orders of magnitude higher than the speed of sound, there are fundamental implications of time synchronization and propagation delays for localization. Additionally, existing communication protocols are not designed to deal with long sleep times and they can't shut down and quickly restart.

In wireless sensor and micro-sensor networks, energy consumption is a key factor for the sensor lifetime and accuracy of information. Protocols and mechanisms have been proposed for energy optimization considering various communication factors and types of applications. Conserving energy and optimizing energy consumption are challenges in wireless sensor networks, requiring energy-adaptive protocols, self-organization, and balanced forwarding mechanisms.

We take here the opportunity to warmly thank all the members of the SENSORCOMM 2013 Technical Program Committee, as well as the numerous reviewers. The creation of such a high quality conference program would not have been possible without their involvement. We also kindly thank all the authors who dedicated much of their time and efforts to contribute to SENSORCOMM 2013. We truly believe that, thanks to all these efforts, the final conference program consisted of top quality contributions.

Also, this event could not have been a reality without the support of many individuals, organizations, and sponsors. We are grateful to the members of the SENSORCOMM 2013

organizing committee for their help in handling the logistics and for their work to make this professional meeting a success.

We hope that SENSORCOMM 2013 was a successful international forum for the exchange of ideas and results between academia and industry and for the promotion of progress in the area of sensor technologies and applications.

We are convinced that the participants found the event useful and communications very open. We hope Barcelona provided a pleasant environment during the conference and everyone saved some time for exploring this beautiful city.

SENSORCOMM 2013 Chairs:

SENSORCOMM Advisory Chairs
Jean Philippe Vasseur, Cisco Systems, Inc., France
Petre Dini, Concordia University, Canada / China Space Agency Center, China
Jaime Lloret Mauri, Polytechnic University of Valencia, Spain
Jens Martin Hovem, Norwegian University of Science and Technology, Norway
Pascal Lorenz, University of Haute Alsace, France
Sergey Yurish, IFSA, Spain

SENSORCOMM 2013 Industry Liaison Chairs
Sarfraz Khokhar, Cisco Systems, Inc., USA
Harkirat Singh, Samsung Electronics Co., Korea
Javier Del Ser Lorente, TECNALIA-Telecom - Zamudio (Bizkaia), Spain
Michael Niedermayer, Fraunhofer IZM, Germany

SENSORCOMM 2013 Research/Industry Chairs
Hristo Djidjev, Los Alamos National Laboratory, USA Teng Rui, National Institute of Information and Communication Technology, Japan
S. Biju Kumar, Philips Research - Eindhoven, The Netherlands

SENSORCOMM 2013 Special Area Chairs
Embedded systems
Joshua Ellul, Imperial College, London, UK
Security
Yenumula Reddy, Grambling State University, USA
Body networks
Alessandro Pozzebo, Università degli Studi di Siena, Italy
Underwater systems
Mylène Toulgoat, Communications Research Centre - Ottawa, Canada
Applications
Elena Gaura, Coventry University, UK
Performance
Canfeng Chen, Nokia Research Center - Beijing, China

Atmospheric Icing and Sensing

Muhammad Shakeel Virk, Narvik University College, Norway

SENSORCOMM 2013

Committee

SENSORCOMM Advisory Chairs

Jean Philippe Vasseur, Cisco Systems, Inc., France
Petre Dini, Concordia University, Canada / China Space Agency Center, China
Jaime Lloret Mauri, Polytechnic University of Valencia, Spain
Jens Martin Hovem, Norwegian University of Science and Technology, Norway
Pascal Lorenz, University of Haute Alsace, France
Sergey Yurish, IFSA, Spain

SENSORCOMM 2013 Industry Liaison Chairs

Sarfraz Khokhar, Cisco Systems, Inc., USA
Harkirat Singh, Samsung Electronics Co., Korea
Javier Del Ser Lorente, TECNALIA-Telecom - Zamudio (Bizkaia), Spain
Michael Niedermayer, Fraunhofer IZM, Germany

SENSORCOMM 2013 Research/Industry Chairs

Hristo Djidjev, Los Alamos National Laboratory, USA Teng Rui, National Institute of Information and Communication Technology, Japan
S. Biju Kumar, Philips Research - Eindhoven, The Netherlands

SENSORCOMM 2013 Special Area Chairs

Embedded systems
Joshua Ellul, Imperial College, London, UK

Security
Yenumula Reddy, Grambling State University, USA

Body networks
Alessandro Pozzebo, Università degli Studi di Siena, Italy

Underwater systems
Mylène Toulgoat, Communications Research Centre - Ottawa, Canada

Applications
Elena Gaura, Coventry University, UK

Performance
Canfeng Chen, Nokia Research Center - Beijing, China

Atmospheric Icing and Sensing
Muhammad Shakeel Virk, Narvik University College, Norway

SENSORCOMM 2013 Technical Program Committee

Saied Abedi, Fujitsu Laboratories of Europe LTD. - Middlesex, UK
Abdalrahman Al-Qubaa, Newcastle University, UK
Mothanna Alkubeily, Université de Technologie de Compiègne, France
Boushra Alkubily, Université de Technologie de Compiègne, France
Tariq Alsboui, Manchester Metropolitan University, UK
Isabelle Augé-Blum, INSA Lyon - Laboratoire CITI -Villeurbanne, France
Reza Azarderakhsh, The University of Western Ontario, Canada
Sebastian Bader, Mid Sweden University, Sweden
Valentina Baljak, National Institute of Informatics & University of Tokyo, Japan
Dominique Barthel, Orange Labs Division R&D - Meylan, France
Novella Bartolini, "Sapienza" University of Rome, Italy
Majid BayanIl Abbasy, Universidad Nacional de Costa Rica, Costa Rica
Rezaul K Begg, Victoria University, Australia
Paolo Bellavista, University of Bologna, Italy
Karabi Biswas, Indian Institute of Technology - Kharagpur, India
Alessandro Bogliolo, University of Urbino, Italy
Lina Brito, University of Madeira, Portugal
Ioannis Broustis, AT&T Labs Research, U.S.A
Tiziana Calamoneri, “La Sapienza” Università di Roma, Italy
Maria-Dolores Cano Baños, Technical University of Cartagena, Spain
Juan Vicente Capella Hernández, Universidad Politécnica de Valencia, Spain
Berta Carballido Villaverde, United Technologies Research Centre, Ireland
Canfeng Chen, Nokia Research Center - Beijing, China
Shu-Ching Chen, Florida International University - Miami, USA
Hugo Coll Ferri, Polytechnic University of Valencia, Spain
Daniel Curiac, "Politehnica" University of Timisoara, Romania
David Cuartielles, Malmö University, Sweden
Debabrata Das, International Institute of Information Technology - Bangalore, India
Danco Davcev, University for Information Science & Technology "St. Paul the Apostle" - Ohrid, Republic of Macedonia
Javier Del Ser Lorente, TECNALIA-Telecom - Zamudio (Bizkaia), Spain
Jerker Delsing, Lulea University of Technology, Sweden
Behnam Dezfouli, University Technology Malaysia (UTM), Malaysia
Vincenzo Di Lecce, Politecnico di Bari, Italy
Mari Carmen Domingo, Barcelona Tech University, Spain
Wan Du, Nanyang Technological University (NTU), Singapore
Juan Carlos Dueñas López, Universidad Politecnica de Madrid, Spain
Sylvain Durand, LIRMM/Université Montpellier II, France
Imad H. Elhajj, American University of Beirut, Lebanon
Joshua Ellul, Imperial College London, UK
Xiang Fei, Coventry University, UK
Sándor Fekete, Braunschweig Institute of. Technology, Germany
Paulo Felisberto, Institut for Systems and Robotics-Lisbon / Universidade do Algarve, Portugal

Gianluigi Ferrari, University of Parma, Italy
Armando Ferro Vázquez, Universidad del País Vasco - Bilbao, Spain
Paul Fortier, University of Massachusetts Dartmouth, USA
Pedro Furtado, University of Coimbra, Portugal
Miguel Garcia Pineda, Polytechnic University of Valencia, Spain
Elena Gaura, Coventry University, UK
Hamid Gharavi, National Institute of Standards and. Technology (NIST) - Gaithersburg, USA
Chris Gniady, University of Arizona, USA
Stephane Grumbach, INRIA, France
Jianlin Guo, Mitsubishi Electric Research Laboratories - Cambridge, USA
Malka N. Halgamuge, The University of Melbourne, Australia
Mohammad Hammoudeh, Manchester Metropolitan University, UK
Vincent Huang, Ericsson Research - Stockholm, Sweden
Muhammad Ali Imran, University of Surrey, U.K.
Abhaya Induruwa, Canterbury Christ Church University, UK
Vasanth Iyer, International Institute of Information Technology, India
Shaghayegh Jaberi, Islamic Azad University - Tehran, Iran
Imad Jawhar, United Arab Emirates University - Al Ain, UAE
Zhen Jiang, West Chester University, USA
Miao Jin, University of Louisiana at Lafayette, U.S.A.
Adrian Kacso, University of Siegen, Germany
Aravind Kailas, University of North Carolina - Charlotte, USA
Kyoung-Don Kang, Binghamton University, USA
Riad Kanan, The Institute of Engineering Sciences, Switzerland
Dimitrios A. Karras, Chalkis Institute of Technology, Hellas
Fotis Kerasiotis, University of Patras / Rio-Patras, Greece
Sarfraz Khokhar, Cisco Systems Inc., USA
Kwangsoo Kim, Electronics and Telecommunications Research Institute (ETRI), Korea
Thorsten Kramp, IBM Research Zurich, Switzerland
Evangelos Kranakis, Carleton University - Ottawa, Canada
Srđjan Krčo, Ericsson Research, Ireland
Dilip Krishnaswamy, Qualcomm Research - San Diego, U.S.A.
Danny Krizanc, Wesleyan University - Middletown, USA
Erlend Larsen, The Norwegian Defence Research Establishment (FFI) - Kjeller, Norway
Seongsoo Lee, Soongsil University - Seoul, Korea
Xiuqi Li, University of North Carolina at Pembroke, U.S.A
Chiu-Kuo Liang, Chung Hua University - Hsinchu, Taiwan
Qilian Liang, University of Texas at Arlington, USA
Weifa Liang, Australian National University - Canberra, Australia
Chun-Lung Lin, Industrial Technology Research Institute (ITRI) Hsinchu, Taiwan
Thomas Lindh, STH/KTH - Stockholm, Sweden
André Luiz Lins de Aquino, Federal University of Ouro Preto, Brazil
Hai Liu, Hong Kong Baptist University, Hong Kong
Jaime Lloret Mauri, Polytechnic University of Valencia, Spain
Elsa María Macías López, University of Las Palmas de Gran Canaria, Spain
Abdallah Makhoul, Université de Besancon - Belfort, France
Vladimir Marbukh, NIST, USA
José Ramiro Martínez de Dios, University of Seville, Spain

Alireza Masoum, Twente University, The Netherlands
Kovatsch Matthias, ETH Zurich, Switzerland
Natarajan Meghanathan, Jackson State University, USA
Fabien Mieyeville, Ecole Centrale Lyon - INL (Institute of Nanotechnology of Lyon), France
Nader Faisal Jaafar Mohamed, UAEU, UAE
Jose M. Moya, Universidad Politécnica de Madrid, Spain
Mohammad Mozumdar, California State University, Long Beach, USA
Abderrahmen Mtibaa, Texas A&M University, U.S.A.
Mughal Umair Najeeb, High North Technology Center - Narvik University College, Norway
Deok Hee Nam, Wilberforce University, USA
Enrico Natalizio, INRIA Lille - Nord Europe, France
Mahmuda Naznin, Bangladesh University of Engineering and Technology - Dhaka, Bangladesh
Sarmistha Neogy, Jadavpur University, India
Edith C.-H. Ngai, Uppsala University, Sweden
Michael Niedermayer, Fraunhofer Institute for Reliability and Microintegration, Germany
Gregory O'Hare, University College Dublin (UCD), Ireland
Brendan O'Flynn, Tyndall National Institute/University College Cork, Ireland
Knut Øvsthus, Bergen University College, Norway
Carlos Enrique Palau Salvador, Universidad Politecnica de Valencia, Spain
Sung-Joon Park, Gangneung-Wonju National University, South Korea
Leonidas Perlepes, University of Thessaly, Greece
Dirk Pesch, Cork institute of Technology, Ireland
Miodrag Potkonjak, University of California - Los Angeles, USA
Shrisha Rao, International Institute of Information Technology - Bangalore, India
Shahid Raza, Swedish Institute of Computer Science (SICS) – Stockholm, Sweden
Yenumula Reddy, Grambling State University, USA
Juan Reig Pascual, Polytechnic University of Valencia, Spain
Càndid Reig, University of Valencia, Spain
Tor Arne Reinen, SINTEF ICT, Norway
Joel Rodrigues, University of Beira Interior, Portugal
Lorenzo Rubio-Arjona, Universidad Politécnica de Valencia, Spain
Teng Rui, National Institute of Information and Communication Technology, Japan
Jorge Sá Silva, University of Coimbra, Portugal
Husnain Saeed, National University of Sciences & Technology (NUST), Pakistan
Addisson Salazar, Polytechnic University of Valencia, Spain
Ioakeim Samaras, Aristotle University of Thessaloniki, Greece
Francisco Javier Sánchez Bolumar, ADIF, Spain
Olga Saukh, ETH Zurich, Switzerland
Kamran Sayrafian, National Institute of Standards & Technology, USA
Leo Selavo, University of Latvia, Latvia
Sandra Sendra Compte, Polytechnic University of Valencia, Spain
Kuei-Ping Shih, Tamkang University - Taipei, Taiwan
Simone Silvestri, Sapienza University of Rome, Italy
Francesco Simeone, University "Sapienza" of Roma / INFN-Roma, Italy
Jasvinder Singh, Cork Institute of Technology, Ireland
K M Sunjiv Soyjaudah, University of Mauritius, Mauritius
Arvind K. Srivastava, NanoSonix Inc. - Skokie, USA
Grigore Stamatescu, University Politehnica of Bucharest, Romania

Yannis Stamatiou, University of Patras, Greece
Razvan Stanica, National Polytechnic Institute of Toulouse, France
Kris Steenhaut, Vrije Universiteit Brussel, Belgium
Julinda Stefa, Sapienza University of Rome, Italy
Ewa Szynkiewicz, Warsaw University of Technology, Poland
Zahra Taghikhaki, University of Twente, the Netherlands
Muhammad Tariq, Waseda University - Tokyo, Japan
Lothar Thiele, ETH Zurich, Switzerland
Rolf Thomasius, Technische Universität Berlin, Germany
Bin Tong, Microsoft Corp. - Redmond, USA
Vassilis Triantafillou, Technological Educational Institution of Messolonghi, Greece
Neeta Trivedi, Aeronautical Development Establishment- Bangalore, India
Wilfried Uhring, University of Strasbourg, France
Fabrice Valois, INRIA SWING / CITI, INSA-Lyon, France
Jean Philippe Vasseur, Cisco Systems, Inc., France
Roberto Verdone, Università di Bologna, Italy
Manuela Vieira, UNINOVA/ISEL, Portugal
Muhammad Shakeel Virk, Narvik University College, Norway
Michael Walsh, Tyndall National Institute, Ireland
You-Chiun Wang, National Chiao-Tung University, Taiwan
Chih-Yu Wen, National Chung Hsing University - Taichung, Taiwan
Andreas Wombacher, University of Twente, Netherlands
Fang-Jing Wu, National Chiao Tung University, Taiwan
Dongfang Yang, National Research Council Canada - London, Canada
Rehana Yasmin, University of Birmingham, UK
Chih-Wei Yi, National Chiao Tung University, Taiwan
Norihiko Yoshida, Saitama University, Japan
Sergey Y. Yurish, IFSA - Barcelona, Spain
Yifeng Zhou, Communications Research Centre, Canada
Tanveer Zia, Charles Sturt University, Australia

Copyright Information

For your reference, this is the text governing the copyright release for material published by IARIA.

The copyright release is a transfer of publication rights, which allows IARIA and its partners to drive the dissemination of the published material. This allows IARIA to give articles increased visibility via distribution, inclusion in libraries, and arrangements for submission to indexes.

I, the undersigned, declare that the article is original, and that I represent the authors of this article in the copyright release matters. If this work has been done as work-for-hire, I have obtained all necessary clearances to execute a copyright release. I hereby irrevocably transfer exclusive copyright for this material to IARIA. I give IARIA permission or reproduce the work in any media format such as, but not limited to, print, digital, or electronic. I give IARIA permission to distribute the materials without restriction to any institutions or individuals. I give IARIA permission to submit the work for inclusion in article repositories as IARIA sees fit.

I, the undersigned, declare that to the best of my knowledge, the article is does not contain libelous or otherwise unlawful contents or invading the right of privacy or infringing on a proprietary right.

Following the copyright release, any circulated version of the article must bear the copyright notice and any header and footer information that IARIA applies to the published article.

IARIA grants royalty-free permission to the authors to disseminate the work, under the above provisions, for any academic, commercial, or industrial use. IARIA grants royalty-free permission to any individuals or institutions to make the article available electronically, online, or in print.

IARIA acknowledges that rights to any algorithm, process, procedure, apparatus, or articles of manufacture remain with the authors and their employers.

I, the undersigned, understand that IARIA will not be liable, in contract, tort (including, without limitation, negligence), pre-contract or other representations (other than fraudulent misrepresentations) or otherwise in connection with the publication of my work.

Exception to the above is made for work-for-hire performed while employed by the government. In that case, copyright to the material remains with the said government. The rightful owners (authors and government entity) grant unlimited and unrestricted permission to IARIA, IARIA's contractors, and IARIA's partners to further distribute the work.

Table of Contents

Techniques for Increasing Network Functionality while Remaining within Legal Maximum TX Duty Cycle Requirements

Eoin O'Connell, Victor Cionca, Brendan O'Flynn
Tyndall National Institute
University College Cork
eoin.oconnell@tyndall.ie
Ireland

***Abstract*—In this paper, a technique is presented which allows users of the license free ISM bands to increase functionality of the wireless network while remaining within the maximum legally allowed transmit duty cycle requirements. We show through analytical modeling and empirical evaluation, that traditional MAC and routing techniques result in a significant increase in the overall TX duty cycle when sensor sample frequencies are increased. Specifically, we focus on the 868MHz ISM band where the maximum allowed legal TX duty cycle must be $< 1\%$. Two simple techniques are presented that significantly reduce the TX duty cycle. The result of this technique is a reduction in power consumption and more importantly, a method to increase network functionality while remaining within the legal requirements. We show that these techniques allow for sensor readings to be collected far more frequently in multi-hop, receiver duty cycled, wireless sensor networks.**

***Keywords*—868MHz limitations, Duty Cycle, Low Power, WSN, Multi-Hop**

I. INTRODUCTION

Considering that the license free bands can be occupied by multiple co-existing wireless networks, transmit power levels and maximum transmit duty cycle restrictions are placed on the users. These sanctions are placed by the ISM band regulatory authorities [1]. Examples of such systems which can be found in residential homes, would be wireless smoke alarm systems and wireless burglar alarm systems. These restrictions are put in place to maximize fairness and reduce the probability of interference/ collisions. Operation in the license free $868-868.6$ MHz ISM band requires maximum TX duty cycles of $<1\%$ (Class 2) and maximum TX power levels of approximately +14dBm. Other nearby bands require $<0.1\%$ (Class 1). This specific duty cycle requirement of $<1\%$ translates to a maximum of 36 seconds of TX activity within one hour and a maximum length of 3.6 seconds for any single transmission. (See Table I).

These restrictions place limitations on the functionality of the network, specifically, they limit the rate at which sensor readings can be reported to the network sink. Increasing the sensor reporting rate increases the TX duty cycle. The rate at which sensor readings must be reported from each node in the network, is a largely application based requirement. Depending on the application, these regulatory restrictions may impose a limit on the desired sensor report rate and the network engineer will need to be satisfied with a lower sampling rate. To overcome the issues described above, two techniques are implemented to reduce the overall TX duty cycle of the multi-hop network as a whole. The first technique and the one that has the largest impact on reducing the TX duty cycle is a neighbor schedule learning system. Using this layer 2 MAC protocol optimization, nodes learn when neighboring nodes are expected to wake up and choose the most energy efficient time to begin contacting them. The result of this is a drastic reduction in TX duty cycle.

TABLE I: Frequency Bands For Non-Specific Short Range Devices in Europe

Frequency Band	Max ERP	TX Duty Cycle	Bandwidth
$868-868.6$MHz	14dBm	$< 1\%$	No Limits
$868.7-869.2$MHz	14dBm	$< 0.1\%$	No Limits

The second technique is a layer 3 routing optimization, whereby nodes piggyback their sensor readings into packets from upstream nodes within the multi-hop network. When a packet is received which requires forwarding towards the network sink, this triggers the recipient of the packet to carry out a sensor reading. This optimization results in fewer transmissions and more optimal utilization of the maximum payload length of the physical layer itself.

A. TX on time during transmissions

The length of time for which the radio is in TX mode during each transmission to a duty cycled neighbor, varies greatly from MAC protocol to protocol. To guarantee reliable communication, non-acknowledge based MAC protocols such as B-MAC and BoX-MAC1 are required to leave the radio transceiver in TX mode for the full duration of the receive check interval (T_W) [8], [6]. The receive check interval is the time interval between receive check operations in duty cycled WSN nodes. Conversely, acknowledge based protocols such as X-MAC and BoX-MAC2 stop transmitting as soon as an ACK is received [2], [6]. On average, this ACK packet will be received half way through the receive check interval. Interestingly, the total length of time for which both of these protocols spend in TX mode during transmissions to duty cycled neighbors, is approximately equal to $\frac{T_W}{4}$. On average, the overall time the radio is active for during transmissions is

 ISBN: 978-1-61208-296-7

indeed $\frac{T_W}{2}$. During this time period of $\frac{T_W}{2}$, the radio spends a share of its time in TX mode sending packets and the rest in RX mode listening for ACK (acknowledge) packets. In the case of BoX-MAC2, the proportions of time spent in RX and TX will vary as the payload length changes. X-MAC on the other hand spends fixed lengths of time in both TX and RX during the contacting phase due to its RTS (Request To Send)/ CTS (Clear To Send) system.

The goal of this work is to provide a simple technique to lower TX duty cycle, allowing users of the license free bands to increase network functionality.In Section II, a detailed description is given into the functionality of both of these techniques. In Section III, modeled results are presented showing the potential savings in TX duty cycle that can be achieved. In Section IV, the experimental setup is explained, empirical test data is presented and compared against our simulated results. In Section IV-D, we discuss these important results and their implications. Finally a brief summary of related work is presented in Section V and in Section VI we conclude and suggest a few topics for further expanding this work.

II. DESIGN

The MAC protocol used for testing is structurally very similar to X-MAC. It consists of an RTS/ CTS preamble wakeup stream which contains source and destination address information. A node which receives an RTS packet addressed to itself, responds with a CTS (containing source and destination address) packet to the sender. Upon reception of the CTS packet at the sender, it proceeds by sending the payload data and waits for a payload acknowledge message. The duration of one send RTS packet and listen for CTS cycle is 2.4ms, one cycle consists of 1.2ms of TX time and 1.2ms of RX time listening for a CTS packet.

A. Reducing TX On Time through Neighbor Schedule Learning

To reduce the energy required to send data packets, a mechanism to learn the wakeup schedules of neighboring nodes is implemented. Learning schedules involves no additional synchronization packets. Each node in the network maintains its own wakeup schedule, and simply guarantees that each periodic receive check will occur at an integer multiple of the network wide receive check interval.

The times at which nodes wake up are random in terms of when the application firmware begins running, but fixed in terms of when they are allowed to perform receive checks. The firmware on each node may begin operating at different times, but the interval between receive checks will remain to be an integer multiple of global receive check period for the duration of the nodes lifetime (excluding drift). Nodes must obey this wakeup schedule independent of their own tasks.

With wakeup scheduling now handled by the aforementioned mechanisms, each node in the network can learn the time offset between its wakeup schedule and those of its neighbors. During the very first interaction between two nodes, the sender stores the number of RTS/CTS cycles it required before the destination responded. This number of *attempts* is stored in a neighbor specific data structure and is proportional to the time offset between their respective wakeup schedules.

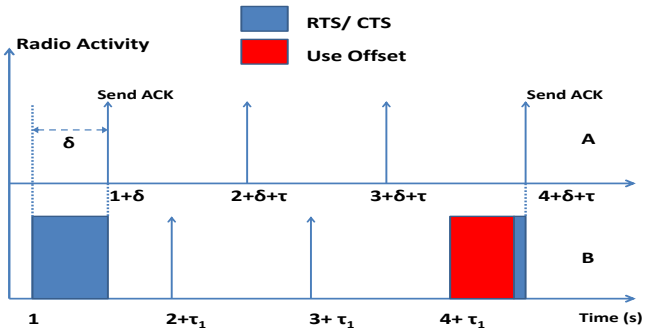


Fig. 1: Vertical arrows indicate when each node wakes at integer multiples of the receive check interval (1 second here). Node B learns the time offset to Node A and applies this learned offset (δ) during next transmission. Graph also illustrates how receive checks will drift (τ,τ_1)

One additional byte is required to store the time offset between wakeup schedules, hence the system is scalable.

During the next transmission to this particular neighboring node, the sender now knows reasonably accurately, when the destination will wakeup. The sender recalls the time offset value from the neighbor specific data structure, and delays accordingly so as to begin performing RTS/CTS cycles a few ms before the destination is expected to wakeup. This time offset is updated during every encounter to account for oscillator drift. The offset learning process is illustrated in Figure 1.In Figure 1, node B first learns the time offset between its wakeup schedule and that of node A, during the next transmission it uses the learned time offset and delays accordingly before contact the destination. In our implementation, the sender begins attempting to contact the destination 9.6ms (T_{SYNC} = 4 RTS/ CTS cycles) before it is expected to wakeup.

B. Piggybacking Data Messages

The primary task of each node in the WSN is to report periodic sensor readings to the network's sink. To reach the network sink, nodes at the outer edges of the deployment may have to route through several nodes, depending on the density and RF environment of the deployment.

The underlying idea of our piggybacking optimization is as follows: nodes that happen to lie in a path that has neighboring nodes generating or forwarding data packets will piggyback their sensor readings into messages which are being forwarded. This process is described graphically in Figure 2. Traditionally, this is done differently and each individual node generates its own periodic data messages.

To accommodate piggybacking, the payload is partitioned into different blocks. Each node which forwards the packet, adds its sensor readings to the payload in a specific position. The position is dependent on the hop number. The leaf node that generated the packet, adds its sensed data to position 0. The next node to interact with and forward the packet adds its sensed data to position 1. In our implementation and application, each node adds a total of 5 bytes to the payload and the length of the variable data packet. Additionally, packets

 ISBN: 978-1-61208-296-7

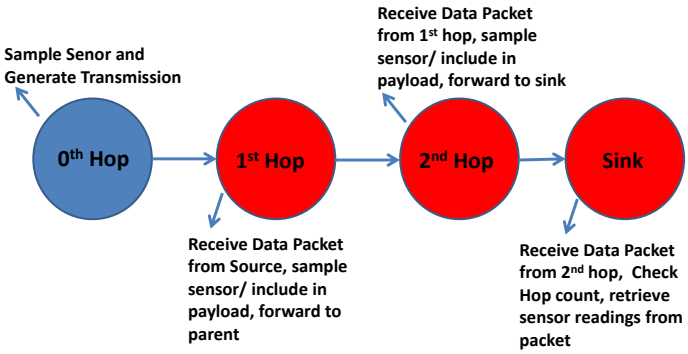


Fig. 2: Data piggybacking. Each node along the path intelligently adds their sensor readings to the payload of incoming messages. Reduces extra wasteful transmissions

contain a routing header. Forwarding nodes increment a *Hop Count* value, and add the ID and RSSI of the last hop to the routing header. Packets that originate from nodes which have a direct RF path to the network sink are short. Payload lengths grow linearly as the hop count increases. At the bit rate of the physical layer in use, each byte adds a total of 80μS to the length of the payload.

When the network sink receives a unique sensed data packet, it first examines the hop count of the packet in the routing header. Depending on this value, it knows how many nodes have interacted with the packet and included sensor readings. It is also aware of the position in the payload where to find sensor readings from the Nth hop.

III. Simulation

To estimate the TX duty cycle across different data send intervals with and without optimizations, the system was modeled in Matlab. The model calculates the overall TX duty cycle resulting in sending data packets at different rates,when one child node is attached and assumes sending to a duty cycled neighbor. The algorithm used is shown in Algorithm 1. Each transmission incurs a total TX time of T_{SYNC}ms worth of RTS/ CTS attempts, i.e., 4.8ms of TX time, plus the length of the payload. A payload length of 24 bytes is 2.4ms at the bit rate of the physical layer. The simulated experimental length was 2 days. The TX duty cycle is hence the total time spent

Algorithm 1 Model of TX on time during transmissions, neighbor learning enabled

```
Initialize Variables
repeat
   Time In TX = 0.0048s + Payload Length
   Accumulated_Time += Send Interval
until Accumulated_Time > 2Days
Duty Cycle = Time In TX / 2 Days
```

in TX mode divided by the length of the test. The results are shown in Figure 3. This graph includes 4 curves, two for neighbor learning enabled and two for neighbor learning disabled. The two curves where neighbor learning is enabled, include piggybacking enabled and disabled. The same applies for the other 2 curves, where neighbor learning is disabled.

TABLE II: Simulated duty cycle values when changing the data send intervals, NL=Neighbor Learn, PB=Piggybacking, None=No optimizations

Interval	NL+PB	NL	PB	None
1s	0.72%	1.44%	25.4%	50.8%
6s	0.12%	0.24%	4.233%	8.47%
20s	0.036%	0.072%	1.27%	2.54%
40s	0.018%	0.036%	0.635%	1.27%

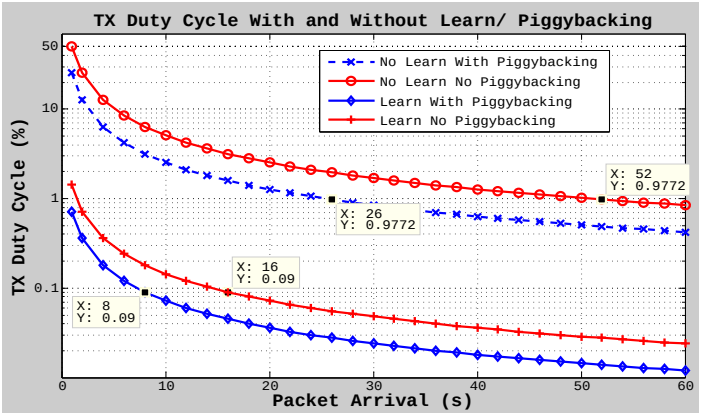


Fig. 3: Results of simulation showing the reduction in TX duty cycle due to Neighbor Learning and Piggybacking

In the case of neighbor learning and piggybacking being enabled, the simulation predicts a 70 times reduction in TX duty cycle compared to both being disabled. With both optimizations, nodes are capable of achieving 0.72% TX duty cycle forwarding data packets every 1 second. With both optimizations disabled this time increases to 52 seconds.The results are summarized in Table II. For each transmission, the node without neighbor learning spends on average half of the receive check interval performing RTS/ CTS, in this scenario that is 0.5s. Of that 0.5s half is spent in TX mode, or 0.25s (as previously discussed in Section II).

In the case of only the piggybacking optimization being enabled, there are two unique scenarios. Firstly, the node under test has piggybacking disabled and it forwards data packets for its child node and also generates its own additional data packets.The simulation shows it is only able to provide $< 1\%$ TX duty cycle when packets are being generated every 52 seconds. When piggybacking alone is enabled, the nodes can report back sensor readings every 26 seconds while remaining $< 1\%$ TX duty cycle.

IV. Empirical Evaluation

A. Testbed

The testbed used for this experiment consists of custom devices, comprised of a PIC24F microcontroller and an SX1211 868MHz radio transceiver. The radio is operated at 100kbps data rate and a sleep current of 1μA is achieved for the platform. The hardware platform is pictured in Figure 4.

 ISBN: 978-1-61208-296-7

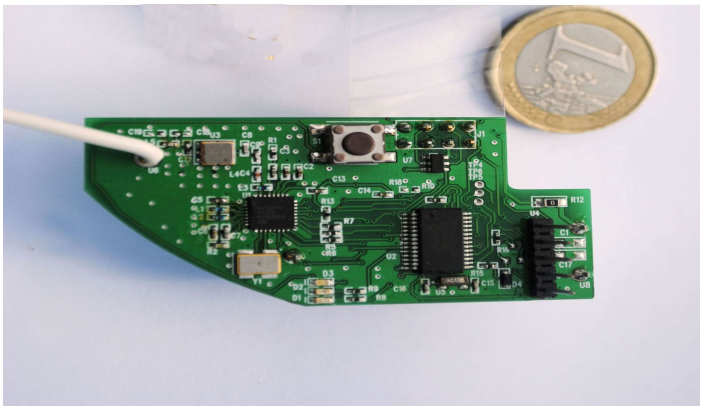


Fig. 4: Testbed WSN node, SX1211 868MHz radio and PIC24F microcontroller

B. Measuring TX on time

To be able to accurately estimate the overall TX duty cycle of deployed nodes, a counter is implemented in software which keeps track of the total amount of time spent in TX mode. This counter is included in each transmission and allows the network sink to keep track of how much time each node spends in TX mode.

The basic unit of time for this counter is 1ms, each increment of the 32-bit counter is equal to 1ms and it overflows after 60 days of constant TX on time. 1ms is the length of an RTS packet and is hence a convenient unit of time. Each transmission which takes place, results in the counter being incremented by the number of RTS/ CTS attempts required to contact the destination, plus the length of the payload transmission. This technique enables very accurate monitoring of the time spent in TX mode at each node. An example of this counter working is a transmission which requires 100 RTS/ CTS attempts to contact the destination and a payload length of 2ms. This transmission would results in the counter being incremented by 102, 100 for the RTS transmissions and 2 for the length of the payload.

The network sink which always listens, sends all received data packets to a PC over USB, where the results are logged. Using the technique described above, the TX duty cycle can be easily calculated for each node using Equation 1.

$$\text{Duty Cycle}(\%) = 100 \times \frac{\text{Packets} \times 0.0011}{\text{Length of Test}} \qquad (1)$$

A screen-shot of the printout from the network sink is shown in Table III. Here nodes 34 and 35 are forwarding packets through node 12 every 5 seconds.The TX duty cycle counter counts the total time spent in TX during transmissions, as can be seen in Table III, the difference between the two consecutive transmissions is 5 counter units or (5*1ms) 5ms. Also included is the RSSI of each hop, -54dBm from nodes 34 and 35 to node 12, and -44dBm from node 12 to the network sink. Most importantly from Table III the packet counter feature can be seen, node 35 requires 5 packets to send its status message (1395-1390).

TABLE III: Log at Network Sink

Hop	ID	Pkts	RSSI	ID	Pkts	RSSI	Time
R1	35	1390	-64	12	7183	-44	7:1:23:45
R1	34	1316	-54	12	7186	-44	7:1:23:48
R1	35	1395	-64	12	7189	-44	7:1:23:50
R1	34	1321	-54	12	7192	-44	7:1:23:53

C. Experimental Setup

To validate our modeled results presented in Section III a small network was deployed using the node pictured in Figure 4. The node's firmware also contains layer 3 routing protocols and these may cause fractional overheads in terms of TX duty cycle. An example of one of these overheads is a periodic probe message to check if the network sink is in range. The network sink was configured to be in an always listening state. This reduces the TX duty cycle of nodes which can communicate with the network sink because they only ever need 1 RTS/ CTS cycle before the payload data can be sent. All nodes were programmed to perform receive checks once a second and forward data packets at the same rate. The sensor sample interval and hence packet generation intervals were chosen to be 1, 2, 5, 10, 20, 30, 40 and 60 seconds.

The first experiment was devised to measure only the difference between nodes with neighbor learning enabled and disabled. For this experiment which was conducted, a total of 17 nodes were deployed in a large office and a maximum of 2 hops was observed. After having deployed the 17 nodes, it was observed that 11 of the nodes were able to communicate directly with the network sink. The remaining 6 were forced to route their messages through the 11 nodes which had a path to the network sink. The nodes of interest were the 6 nodes which did not have a direct RF path to the network sink because they were required to send to duty cycled neighbors. 3 of them were programmed with neighbor learning enabled and 3 without. Tests were conducted for 24 hours and all results were logged on a PC which was connected to the network sink. Of the 6 nodes under test, their TX duty cycles were calculated using Equation 1. 'Packets' represents the software packet counter shown in Table III and explained in Section IV-B.

The second experiment was devised to test the piggybacking and neighbor learning technique combined and a slightly different topology was required. The reason being that in the first experiment the leaf nodes were the only nodes sending to duty cycled neighbors. A controlled 3 hop topology was required because the nodes under test needed to be parent nodes, and still forward to duty cycled neighbors. The topology is depicted in Figure 5 and the nodes under test are the yellow nodes (2^{nd} row from right).

D. Results

The first experiment was devised to solely measure the efficiency of our implemented neighbor schedule learning optimization. The TX duty cycle of a total of 6 leaf nodes was measured over a 1 day period (3 with neighbor learning, 3 without). The results are summarized in Table IV. Each test (i.e., each send interval) was carried out a total of 3 times and the maximum,minimum and overall average TX

 ISBN: 978-1-61208-296-7

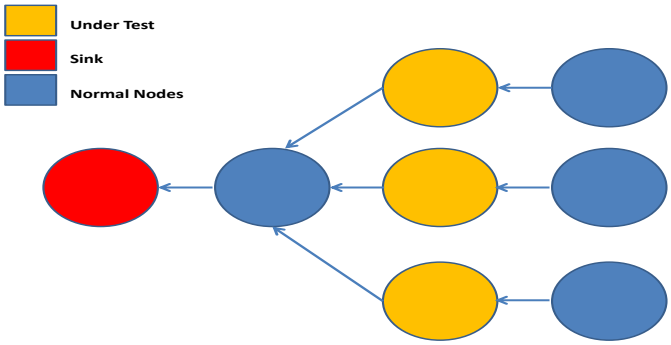


Fig. 5: Topology for 2^{nd} Experiment, Yellow nodes are the nodes under test

duty cycles are listed in Table IV. In Figure 6 , simulated and tested results are compared. The empirically tested values are in extremely close agreement with the modeled results. The maximum variation in the predicted vs the measured duty cycle is 0.044% when sending packets every 40 seconds. Minute variations can be observed between the predicted and tested results, these variations can be explained by layer 3 routing protocol overheads. These variations can be explained by the overheads mentioned in Section IV-C.

The second experiment was designed to measure the efficiency of our piggybacking and neighbor learning algorithm combined. The TX duty cycle of 3 parent nodes each with 1 child node a piece, was measured. The experiment was carried out using a number of different configurations. Different combinations of piggybacking/ neighbor learning enabled and disabled were used. The results are presented in Figure 6. In the case of piggybacking only being disabled, the parent node under test forwards packets for its child and generates its own packet during each data send interval. In the experiment where piggybacking was enabled, the parent node sends only one packet during each data send interval. The difference in TX duty cycle is intuitively approximately 50%, as the workload of the parent is reduced by a factor of 2 when piggybacking is enabled. The slight increase in the payload length due to piggybacking is insignificant when compared to the amount of time spent in TX mode during the RTS/ CTS phase of the transmission (each byte adds only 80μs compared to 1ms for an RTS/ CTS cycle).

In the case of neighbor learning and piggybacking being enabled we observe a 70 times reduction in TX duty cycle, 0.7% forwarding every second (6b), when both are disabled this forwarding rate must be increased to almost 50 seconds to achieve $< 1\%$ TX duty cycle (6a). With both optimizations enabled, parents with a single attached child node can forward packets every 8 seconds to comply with the $< 0.1\%$ TX duty cycle restrictions (6b), with 2 child nodes attached this figure increases to 16 seconds.

From Figure 6a, there is very close correlation between experimental and predicted results. With only piggybacking enabled the data send interval can be 26 seconds while still remaining below 1%, with it disabled, the data send interval must be 52 seconds to comply with the 1% TX duty cycle restrictions.

TABLE IV: Simulated Duty Cycle Values vs Tested, Experiment 1. Minimum, Maximum and Average values included

Interval	Simulated	Test (Avg)	Test (Min)	Test (Max)
1s	0.72%	0.74%	0.71%	0.76%
2s	0.36%	0.37%	0.33%	0.39%
10s	0.072%	0.079%	0.076%	0.078%
20s	0.036%	0.039%	0.036%	0.042%
30s	0.024%	0.026%	0.025%	0.028%

V. Related Work

In terms of the two optimizations and their novelty, there are some similar concepts to be found in the literature. Protocols such as WiseMAC [4] and Hui and Culler's techniques in [5], use similar neighbor schedule learning systems. The differences between WiseMAC and this work are the following: WiseMAC uses an uninterrupted preamble wakeup sequence that doest not contain address information. This work uses an RTS/CTS system which will stop as soon as the destination responds (zero over-listening). WiseMAC requires periodic exchange of scheduling information, this work does not. WiseMAC relies on a Layer 1 receive check, this work uses Layer 2 (i.e., lower power in dense networks due to less overhearing but a slight increase in receive check energy). WiseMAC is also for infrastructure networks and it only considers down-link traffic (parent to children).

In [5], the authors briefly mention and describe their neighbor schedule learning system. They improve upon WiseMAC's preamble only wakeup stream by adding some address information into the wakeup stream. Receive checks are layer 1 based and nodes which overhear can quickly decide the packet is not destined for them because of the embedded address information. But still overhearing does occur, unlike this work. Schedule information is exchanged by including extra data in acknowledge messages, in this work no extra data is transferred to provide neighbor schedule learning.

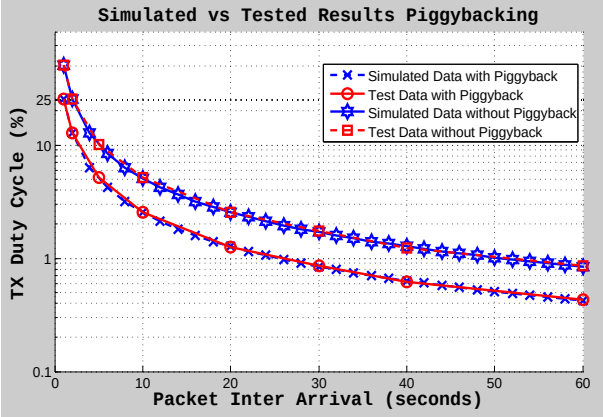
Ye et al. in SCP-MAC [10], present a MAC protocol where all nodes are scheduled to wake during the same time window. Transmissions now take place within this window, resulting in a reduced TX duty cycle compared to standard duty cycled MAC techniques. SCP-MAC requires additional scheduling packets to be transmitted and this has an impact on the overall TX duty cycle, it also suffers from high receive check energy and high latency.

The authors in [4], [5] and [10] present results on the reduction in power consumption of their techniques but fail to investigate the potential reduction in TX duty cycle. This work leverages some of these concepts and simplifies/ optimizes them and applies them to industrial applications.

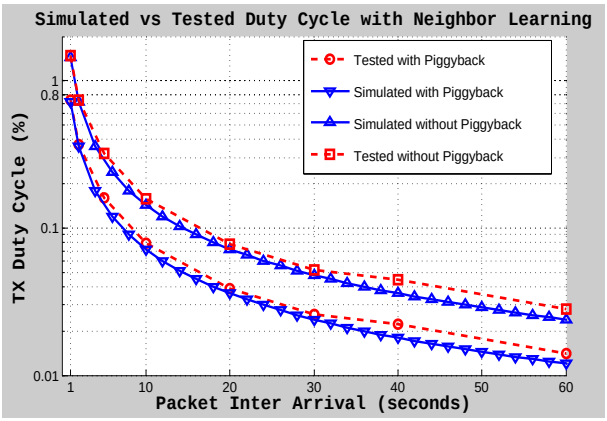
VI. Conclusion & Future Work

In this paper, two techniques to reduce TX duty cycle are described, simulated and empirically evaluated. WSNs which must report real time sensor readings are considered and data aggregation techniques such as in [9] are not considered. Our techniques improve functionality of WSN deployments by allowing users of license free bands to increase network activity, while still remaining within the legal maximum TX

 ISBN: 978-1-61208-296-7



(a) Results of simulation showing the reduction in TX duty cycle due to Piggybacking alone. Upper two curves have piggybacking disabled and lower two have it enabled. Excellent correlation between predicted and tested results. In general piggybacking reduces TX duty cycle by a factor of 2



(b) Neighbor Learning Simulated vs Test Data, excellent correlation between simulated and tested results, maximum error of 0.044% when packets are being sent every 40 seconds. Upper two curves have piggybacking disabled and lower two have it enabled

Fig. 6: Empirical Test Results vs Simulated Results

duty cycle requirements. Using both of the aforementioned techniques, it is demonstrated that sparse (non-dense) networks can provide sensor readings 70 times more frequently to comply with the $< 0.1\%$ TX duty cycle requirements of some of the ISM bands.

Specifically, it is shown that using both techniques and having only 1 dependent child, sensor readings can be forwarded every 8 seconds while still complying, and every 16 seconds when 2 child nodes are attached. The result is a far more active network which is able to provide more frequent sensor readings to the end user. Another important by-product of the drastic reduction in TX duty cycle is the reduction in power consumption and increased lifetime of the battery powered network.

There are a few distinct areas where our work can be improved upon and developed further. For applications where latency is not an issue, large savings in TX duty cycle could be made if nodes which are under a large workload (i.e., multiple child nodes) could queue received data packets and transmit them all in one burst. This would prevent the TX duty cycle from increasing with workload, the disadvantages of this approach would be that sensor readings would no longer be real-time and payload lengths would increase drastically. The performance of the developed techniques in large scale multi-hop deployments is also of interest. The authors would also like to compare their work against other cross layer approaches such as Dozer and Koala [3], [7].

ACKNOWLEDGMENT

The authors would like to acknowledge and thank the support of EI Electronics and IRCSET (Irish Research Council for Science Enterprise and Trade).

REFERENCES

[1] ETSI EN300 220-1, 2012.

[2] Michael Buettner, Gary V Yee, Eric Anderson, and Richard Han. X-mac: a short preamble mac protocol for duty-cycled wireless sensor networks. In *Proceedings of the 4th international conference on Embedded networked sensor systems*, pages 307–320. ACM, 2006.

[3] Nicolas Burri, Pascal von Rickenbach, and Roger Wattenhofer. Dozer: ultra-low power data gathering in sensor networks. In *Proceedings of the 6th international conference on Information processing in sensor networks*. ACM, 2007.

[4] Amre El-Hoiydi and Jean-Dominique Decotignie. Wisemac: An ultra low power mac protocol for multi-hop wireless sensor networks. In SotirisE. Nikoletseas and JosD.P. Rolim, editors, *Algorithmic Aspects of Wireless Sensor Networks*, volume 3121 of *Lecture Notes in Computer Science*, pages 18–31. Springer Berlin Heidelberg, 2004.

[5] Jonathan W. Hui and David E. Culler. Ip is dead, long live ip for wireless sensor networks. In *Proceedings of the 6th ACM conference on Embedded network sensor systems*, pages 15–28, Raleigh, NC, USA, 2008. ACM.

[6] D. Moss and P. Levis. Exploiting physical and link layer boundaries in low-power networking. Technical report, Stanford, 2008.

[7] Razvan Musaloiu-E., Chieh-Jan Mike Liang, and Andreas Terzis. Koala: Ultra-low power data retrieval in wireless sensor networks. In *Proceedings of the 7th international conference on Information processing in sensor networks*. IEEE Computer Society, 2008.

[8] Joseph Polastre, Jason Hill, and David Culler. Versatile low power media access for wireless sensor networks. In *Proceedings of the 2nd international conference on Embedded networked sensor systems*, SenSys '04, pages 95–107, New York, NY, USA, 2004. ACM.

[9] S. Sivaranjani, S. Radhakrishnan, and C. Thangaraj. Adaptive delay and energy aware data aggregation technique in wireless sensor networks. In VinuV Das and Yogesh Chaba, editors, *Communications in Computer and Information Science*, volume 296, pages 41–49–. Springer Berlin Heidelberg, 2013.

[10] Wei Ye, Fabio Silva, and John Heidemann. Ultra-low duty cycle mac with scheduled channel polling. In *Proceedings of the 4th international conference on Embedded networked sensor systems*, pages 321–334, Boulder, Colorado, USA, 2006. ACM.

 ISBN: 978-1-61208-296-7

Unidirectional Link Triangle Routing for Wireless Sensor Networks

Reinhardt Karnapke and Jörg Nolte
Distributed Systems/Operating Systems group
Brandenburg University of Technology Cottbus
Cottbus, Germany
Email: {Karnapke, Jon}@informatik.tu-cottbus.de

***Abstract*—Experiments with wireless sensor networks have shown that asymmetric and unidirectional links do not only exist, but are indeed quite common. Still, many people argue that the gain in connectivity is not worth the effort of making them usable for routing protocols. In this paper, we follow the opposite approach and introduce Unidirectional Link Triangle Routing, which reduces the overhead and, therefore, makes unidirectional links usable on the routing layer.**

Keywords: Wireless Sensor Networks; Routing Protocols; Unidirectional Links

I. Introduction

Unidirectional links represent a serious problem for many routing protocols. Even though experiments with wireless senor networks have shown that they are quite common (e.g. [1], [2], [3], [4]), most routing protocols still ignore their existence or try to remove their implications by using blacklisting or similar methods.

Most of the available protocols that use unidirectional links need to inform the upstream node of its outgoing unidirectional link explicitly, which is often done proactively and introduces a lot of overhead [5]. One possible conclusion that is often drawn from this fact is that using unidirectional links in a routing protocol does not pay off. An alternative is to reduce the overhead produced by the protocols. We argue that the high number of unidirectional links found in experiments makes using them absolutely mandatory and introduce Unidirectional Link Triangle Routing, a routing protocol for wireless sensor networks with often occurring unidirectional links.

The mechanisms used by Unidirectional Link Triangle Routing to make unidirectional links usable without introducing too much overhead are described in Section II, followed by an evaluation both with simulations and real world experiments in Section III. Related work is shown in Section IV. We finish with a conclusion in Section V.

II. Unidirectional Link Triangle Routing

Unidirectional Link Triangle Routing (ULTR) has been designed to use unidirectional links instead of ignoring them or removing their implications. To make them usable, it can cooperate with a neighborhood discovery protocol if one is necessary for the application, or with the MAC layer if a TDMA protocol is used.

To make an existing neighborhood table usable for ULTR, a neighborhood table entry on node A should consist of at least three parts:

1) the ID of the neighbor (e.g. B),
2) the status of the link to that neighbor
3) the ID of a potential forwarding node

The status of a link is either bidirectional, unidirectional-incoming or unidirectional-outgoing. The forwarding node is only necessary if the link is unidirectional-incoming.

When a node wants to transmit a message to another node that is not included in its neighbor table or its routing table, it starts a route discovery by transmitting a route request (RREQ) message. This message is flooded through the network and creates routing entries for the source on all nodes it passes. The entries include only the next hop and the distance, resulting in a distance-vector protocol like AODV [6].

The differences start once the destination has been reached and transmitted the route reply. When a node receives a message that is not flooded, i.e., a route reply (RREP) or DATA message, it checks its routing table to find out which of its neighbors is the intended next hop, just like in AODV. Unlike AODV, there is another step after that one. Once the node knows the neighbor that has been chosen to forward the message, it checks its neighbor table to see if the link to that node is *currently* an unidirectional-incoming one. If it is, and a detour of one hop is possible, the node forwards the packet first to the detour node which, in turn, retransmits the message to the intended node. Otherwise, the message is silently discarded. Please note that broken links may be treated just like unidirectional-incoming ones.

Figure 1 shows a small part of a network and the corresponding neighborhood table entries used in this protocol: The nodes A, B and C are connected bidirectionally, with the exception of the link between nodes A and B, which is unidirectional and enables only transmissions from B to A. The neighborhood table of node A consists of two entries: one bidirectional entry for node C and a unidirectional-incoming one from node B, with node C denoted as designated forwarder. The neighborhood table of node B contains node A, which would not be possible without a two-hop neighborhood discovery protocol, as node B does not receive any messages from node A. The link is marked as unidirectional-outgoing, and, thus, does not need any forwarder. The second entry features node C with a bidirectional link, needing no forwarder either.

 ISBN: 978-1-61208-296-7

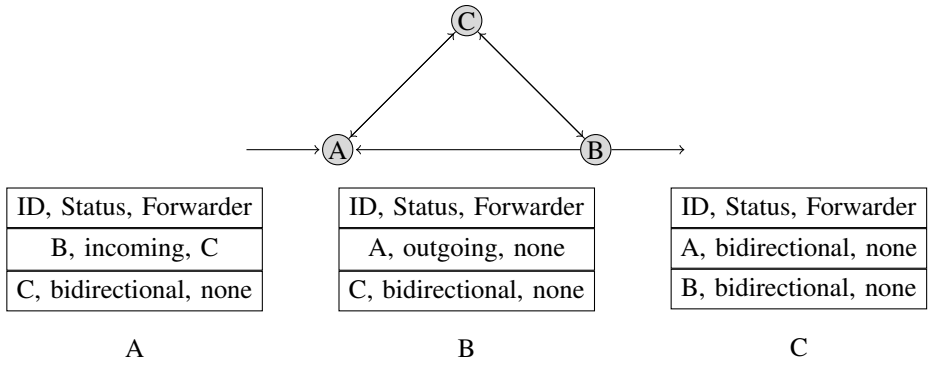


Fig. 1: Neighbor Table Entries in Unidirectional Link Triangle Routing

Finally, the neighborhood table of node C contains nodes A and B, both marked as connected through bidirectional links and not needing any forwarders.

The unidirectional link and the detour that is taken on the way back form a triangle. Therefore this protocol is called Unidirectional Link Triangle Routing.

ULTR is similar to the link layer tunneling mechanism proposed by the unidirectional link working group of the IETF [7], but does not require multiple interfaces on the nodes to communicate. Also, depending on the used neighborhood discovery protocol, it may even be able to work with triangles that include more than one unidirectional link, which the link layer tunneling mechanism cannot handle. Moreover, ULTR works completely on the routing layer, the link layer is not involved. This is an advantage when timeouts are used, because the extra hop and the resulting longer delay are not hidden from the routing layer.

A. Neighborhood Discovery

The neighborhood discovery protocol needed for ULTR may be quite simple and needs only be started on a node once it receives the first message from a neighbor, i.e., when the first route request message is flooded into the network. Once it has been started, the neighborhood discovery protocol should regularly transmit a message containing the IDs of all nodes from which this node has recently received messages and the status of its links to and from them. When a node receives such a hello message, it checks whether its ID is contained therein. If it is not, the receiving node knows that it is on the receiving side of a unidirectional link.

In protocols that do not use unidirectional links, a lot of overhead would now be necessary to inform the upstream node (the sender of the hello message) of the unidirectional link. In this protocol, the upstream node does not need to know about its existence. The receiving node only marks the link as unidirectional-incoming in its neighbor table.

When a node A receives a hello message via the bidirectional link from node C in which the upstream node of the unidirectional link is listed and the link to that node (from C to B) is marked as bidirectional, node A enters the sender of the hello message (node C) as a forwarding neighbor into the corresponding neighbor table entry (for node B). Please note that this would also be possible if there was a unidirectional link from C to B, but the proactive detection would introduce a large overhead and solve only one special case: If there is a unidirectional link from C to B and no other neighbor of A has a bidirectional link to B.

When a message (RREP or DATA) is sent the reverse way, it needs to be forwarded along a one-hop-detour. This message may be used to inform the upstream node of the link, which is then entered into the upstream node's neighborhood table as unidirectional-outgoing. Please note that for the routing alone this information would not be necessary, indeed it would be easy to hide the fact that the message has taken a detour. But for the sake of timers that may be used for retries on MAC, routing or transport layer it helps to know that the delay could be twice as high. In this case the information about this special link may be acquired "for free" and could be used to solve the problem described above. The information about the unidirectional-outgoing link may also be used by the MAC layer not only for retries, but also to determine the correct two-hop neighborhood of a node, which is a mandatory information for TDMA protocols.

B. Message Types

ULTR uses three message types: Route Request, Route Reply and DATA. Figure 2 shows an example for each of them:

A Route Request message contains the identity and sequence number of the source which are used for duplicate detection, followed by the identity of the destination. The hop count is incremented by one on each hop as usual, and the identity of the last hop is used to build the backward route. A Route Reply message contains sequence number and identity of the source for duplicate detection as well as the identity of the destination. For forwarding purposes the next hop and, if necessary, the forwarding node are included. The DATA packet contains the sequence number and identity of its source as well as the identity of its destination. This is followed once again by the identities of the next hop and, if suitable, the forwarding node. The last part of the DATA message contains the application data.

 ISBN: 978-1-61208-296-7

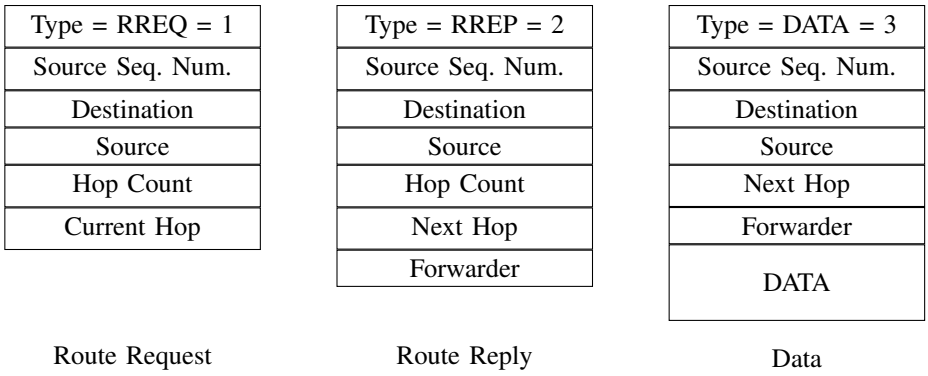


Fig. 2: Message Types Used in ULTR

C. ULTR without Neighborhood Discovery Protocol

ULTR relies on a neighborhood discovery protocol, which supplies information about incoming and outgoing unidirectional links. If neither the application nor the MAC protocol needs a neighborhood discovery protocol, a variation of ULTR with passive link detection may be used. But passive link detection means that sometimes a node does not know about links to its neighbors, even though they are available. Therefore, a second mode of operation is introduced: if a node does not have a link to the next hop in its neighbor table, it forwards the message nonetheless, with an additional flag telling its neighbors that any of them that do have an active link to the next-but-one hop (i.e., the siblings of the next hop) should also forward the message.

When this variation is used, some modifications to the message types are necessary (see Figure 3). Information about the last hop needs to be included in RREQ messages, in addition to the current hop. Both node IDs are stored in the routing table. A node decides which entry to use depending on the overheard status of the link. If the next hop is assumed to be connected by a bidirectional link, the normal next hop is used. Otherwise the message is set to detour mode and the next-but-one hop is used. The last hop is also used for implicit link detection: If a node overhears the transmission of a message in which it is denoted as last hop, it knows that the link between itself and the current hop denoted in the message is *currently* bidirectional.

A RREP message contains three node IDs instead of only two: The last hop ID and current hop ID are used to build the backward route for normal and for detour mode just as they are used in the RREQ. The next hop ID is used for forwarding. However, the RREP also contains a flag denoting the mode of transmission, which can take on the values ”normal” and ”detour”. It is evaluated upon message reception to decide whether a node shall forward the message or not. In normal mode it only forwards the message when it is denoted as the next hop in the message, in detour mode it also forwards the message if it has the next-but-one hop in its neighbor table.

The DATA message features the same three node IDs that are present in the RREP message. For routing purposes alone, the last hop ID would not be needed, but it is nevertheless included for link status detection. The mode flag is also present again, to enable the usage of a one-hop detour if the status of the next link is unknown or known to be unidirectional-incoming.

D. Cooperation with the MAC-Layer

Like all routing protocols that use unidirectional links, ULTR also needs a MAC that can transmit over unidirectional links. The information about the existence of the unidirectional links probably needs to be collected to a certain extend anyway, depending on the MAC protocol used. So either this may be retrieved from the MAC without additional cost, or the MAC protocol can query the routing layer for link information using an appropriate interface.

ULTR was designed specifically to use unidirectional links. This makes it imperative to use a MAC layer that can also transmit over unidirectional links. Any protocol that uses the standard ”request to send” - ”clear to send” mechanism is completely unsuitable, as no ”clear to send” message will ever be received over an outgoing unidirectional link. Moreover, nodes with an outgoing unidirectional link will never know that they could be disturbing the communication between two other nodes. There are some improvements that allow contention based protocols to work with unidirectional links, e.g., ECTS-MAC [8].

Some of the MAC protocols that use unidirectional links route their link layer acknowledgments back to the upstream nodes. For this, the neighborhood table used by ULTR could be reused.

Plan based MAC protocols need to know the two-hop neighborhood of each node to identify the collision domain. Within this domain, the varying parameter (e.g. frequency (FDMA), code (CDMA) or slot (TDMA)) needs to be unique for each node. Therefore, a neighborhood discovery protocol, which finds these two-hop neighbors, is needed. If the MAC protocol already has its own neighborhood discovery protocol, it only needs to make the gathered information available to ULTR.

The usage of such a neighborhood discovery protocol would also implicitly solve the ”special case” of a unidirec-

 ISBN: 978-1-61208-296-7

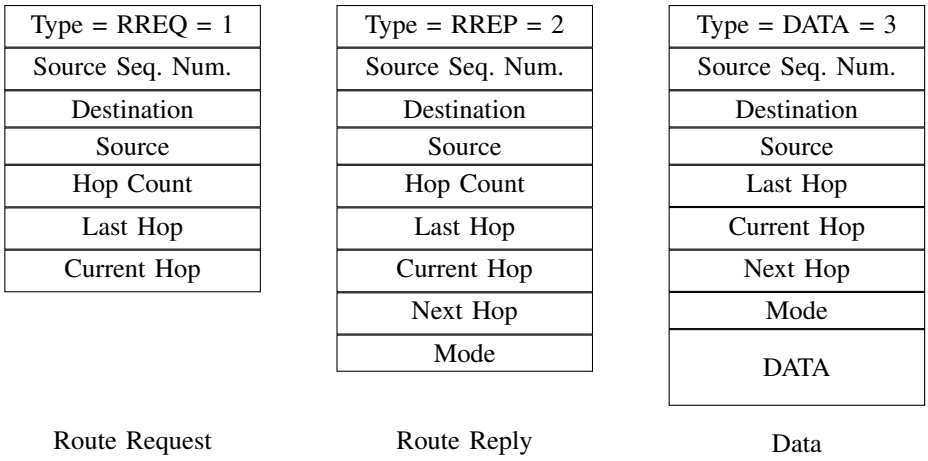


Fig. 3: Message Types in ULTR Without Neighborhood Discovery Protocol

tional link triangle with more than one unidirectional link, enabling ULTR to make use of such links as well.

This usage of a single neighborhood discovery protocol for both MAC and routing reduces communication overhead and memory consumption by far. It also ensures that both layers work on the same data. If they would use different algorithms, different storage sizes or replacement strategies, lots of problems could result, as described, e.g., in Murphy Loves Potatoes [9].

III. Evaluation

The performance of ULTR in the original form depends mainly on the quality of the link information supplied by the neighborhood discovery protocol used by the application. As it is not foreseeable which protocols will be used, the version of ULTR that uses its own passive link detection mechanism was evaluated and compared to four related work protocols: AODVBR [10], DSR [11], Tree Routing and Flooding. Tree Routing was chosen because it is still the most common routing protocol used in wireless sensor networks. DSR was chosen for its ability to handle unidirectional links, and AODVBR has an interesting way of recovering lost data messages. Flooding was also included to define an upper limit to the number of messages that may be delivered in the simulations, where the delivery ratio of a protocol is defined as the number of messages delivered by that protocol divided by the number of messages delivered by Flooding. Determining the maximum number of messages that may be delivered by Flooding was necessary, because the used connectivity model changes links between nodes often and does not guarantee that a path between two nodes exists at all.

The evaluation was split into two parts, simulations and real world experiments. The simulations were based on the discrete event simulator OMNeT++ [12] with the MiXiM [13] extension and were used to evaluate the performance of the selected routing protocols in the presence of unidirectional links, without interference from the MAC layer. Excluding the MAC layer kept the results interpretable.

As the influence of the real hardware and the MAC layer may be quite strong in a sensor network, real word experiments formed the second part of the evaluation. The real word experiments were realized using 36 sensor nodes of type eZ430 Chronos [14] from Texas Instruments and used the CCA MAC supplied by the hardware.

A. Simulation Results

The networks used in the simulations consisted of four sizes of grids: 100 nodes (10×10), 400 nodes (20×20), 900 nodes (30×30) and 1600 nodes (40×40). The grid layout was chosen to represent an application that needs area coverage, with each sensor node placed one distance unit from its direct neighbors above, below, to the right and to the left. However, as mentioned before, the connectivity between nodes was not simply determined by their distance, but rather by a certain probability, that depended only partially on the distance. Also, links changed often and unidirectional links were common in the simulations, as they have been shown to be in the real world. The results presented here are averages of more than 600 simulations for each network size.

The delivery ratio achieved by each protocol in the simulations is shown in Figure 4. It may be seen that the delivery ratio of the related work protocols strongly declines with increasing number of nodes and therefore network diameter. For Tree Routing, this may be explained by the absence of a real route maintenance mechanism. When the forwarding of a message fails due to link break or because the link turned unidirectional incoming, two retransmissions are tried. If the link becomes available again during those two retries, the message has gained one hop, otherwise it is lost. AODVBR uses an intelligent way to recover lost data messages, building a fish bone structure during route discovery. However, this mechanism is used only for data messages, once the initial route has been established. The main problem of AODVBR is this establishment of the initial route, which consists of a route request (RREQ) flooding and a route reply (RREP) transmission along the inverted path taken by the fastest RREQ. If the RREP is lost, e.g., due to the presence of a

 ISBN: 978-1-61208-296-7

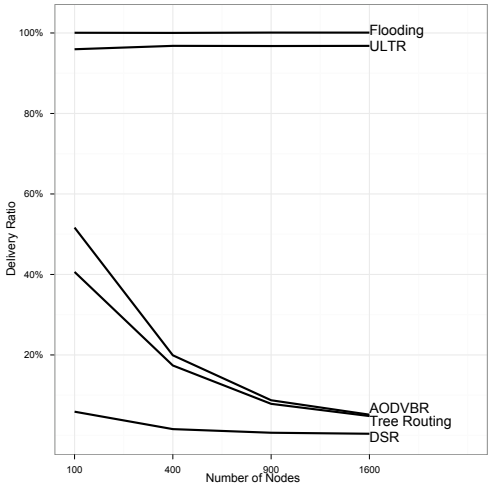


Fig. 4: Delivery Ratio achieved in the Simulations

unidirectional link on the path taken by the RREQ, no initial route may be found. This problem arises quite often due to the nature of unidirectional links, namely their longer reach which often exceeds that of bidirectional links by far. DSR can work in the presence of unidirectional links, which is one of the reasons it was chosen for comparison. However, it suffers heavily from dynamic link changes, as the route maintenance mechanism of DSR in the mode that uses unidirectional links produces a lot of overhead. ULTR on the other side delivers more than 95% of the number of messages delivered by the reference protocol Flooding.

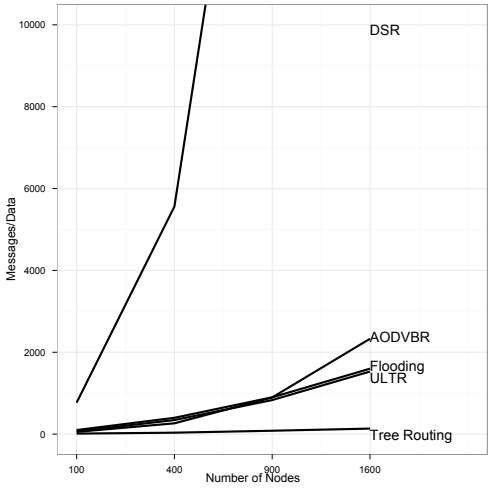


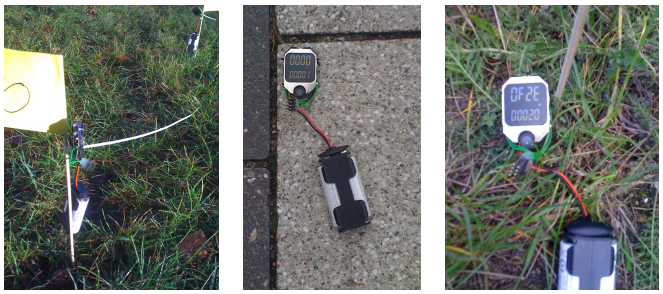
Fig. 5: Number of Messages transmitted to deliver a single Application Message in the Simulations

The efficiency of the protocols measured in messages transmitted in order to deliver a single application message is shown in figure 5. The low number of delivered messages and extremely high number of transmitted messages absolutely disqualify DSR for the simulated type of network with its often changing links. Interestingly, Tree Routing shows the best performance when only the network load is considered. This is due to its simplicity and the low cost of transmission failure: Where other protocols use complex mechanisms to repair the broken route, Tree Routing only uses its two retransmissions, resulting in a maximum cost of 3 times n transmissions, where n is the route length. For comparison: An unsuccessful forwarding of a message in DSR results in a route error message being transmitted to the originator of the message, which then starts a new route discovery by flooding the network with RREQ messages. Once one of these reaches the destination, the network is flooded again with the route reply, due to the operation in unidirectional link mode.

When considering only this figure, Tree Routing seems to be the optimal choice. However, as shown earlier, it has a very low delivery ratio for larger networks. Therefore, it should only be used in networks with a small diameter and when the network load is more important than the delivery of all messages.

B. Real World Experiments

For the real world experiments, 36 sensor nodes were placed in a grid of 6 times 6 nodes in four different locations: On a desktop, affixed to poles, placed on a lawn and placed onto a stone pavement. The transmission power was set to 0dBm.



(a) affixed to poles (b) on a stone pavement (c) placed on the lawn

Fig. 6: A modified eZ430-Chronos Sensor Node

The desktop placement is a one-hop environment, where each node was able to communicate directly with each other one. In the pole placement (figure 6(a)) the nodes were fixed to poles using cable binding, at a height of approximately 20 centimeters above ground with a distance of one meter between nodes, resulting in route lengths between 1 and 2 hops. The stone (figure 6(b)) and lawn (figure 6(c)) placements also used a distance of one meter between nodes, but resulted in route length of up to 5 hops due to the shorter reach of nodes placed on the ground.

Figure 7 shows the delivery ratio of each protocol for the real world experiments, sorted by placement. In the pole placement, AODVBR, DSR and Tree Routing achieved roughly the same delivery ratio. Flooding performed worse due to the high number of messages generated and the resulting MAC layer problems. ULTR performed worst with a delivery ratio of 59%. This is due to problems with the passive link detection, which does not work well in networks with a low diameter. Please note that in the real world experiments 2100 application

 ISBN: 978-1-61208-296-7

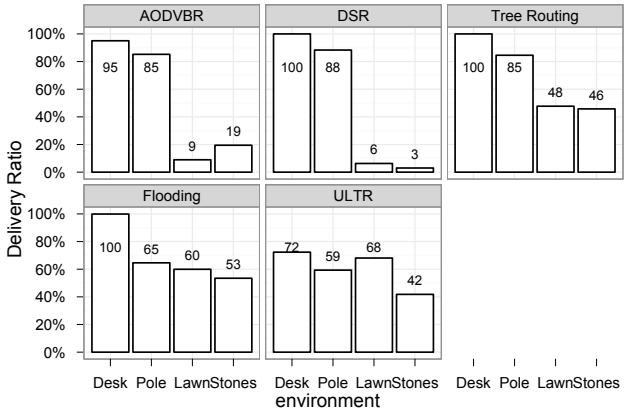


Fig. 7: Delivery Ratio achieved in the Real World Experiments

messages were generated and the delivery ratio was defined as the number of messages delivered by a protocol divided by 2100.

In the lawn and stone experiments, the influence of the route maintenance was stronger, as routes broke more often due to the increase of the average route length. It may be seen once more that the route maintenance mechanism of DSR is not usable in highly dynamic scenarios. AODVBR suffers from the higher probability of having a unidirectional link within its initial routes and Flooding produced too many messages, leading to many collisions. Tree Routing delivered 46 to 48 percent of messages, which may be explained by the node topology. Nodes close to the sink, i.e., those within two hops, were able to deliver their messages most of the time due to the two transmission retries. Those further off were able to deliver only seldom. ULTR performs best in the lawn placement, and nearly equal to Tree Routing in the stone experiments.

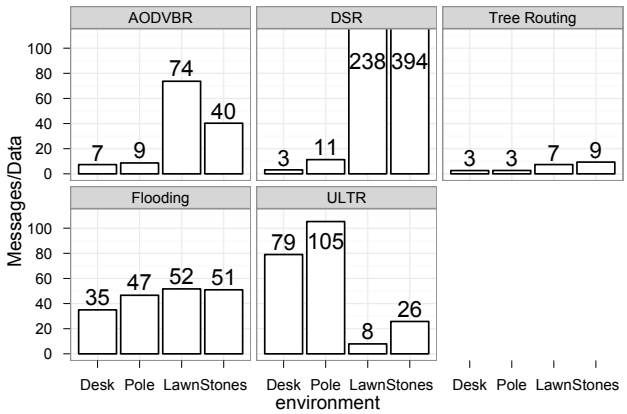


Fig. 8: Number of messages transmitted to deliver a single application message in the Real World Experiments

The cost of delivering an application message measured in transmitted messages is shown in figure 8. As seen in the simulations, Tree Routing is very efficient when the number of messages is considered. This fact is also apparent in all real world experiment placements. When the network load is the limiting factor, Tree Routing should therefore be used, even though it does not reach 50% delivery ratio in the lawn and stone experiments. The other end of the spectrum may be seen in DSR, which transmits 283 and 394 messages per delivered data message in the lawn and stone scenarios respectively. ULTR in its current version should not be used for single hop or 1 - 2 hop scenarios, as the passive neighborhood detection only starts to work in larger networks. There, the performance of ULTR is better than that of all related work protocols, except for Tree Routing. In the evaluated scenarios, ULTR would be the protocol of choice for the lawn placement, and Tree Routing should be used on the stones. The simulations have shown, however, that Tree Routing runs into strong problems when the number of nodes increases, meaning that for larger networks ULTR should be used in both placements, as it scales much better with the network diameter.

C. Importance of Timeouts

The implemented version of ULTR with passive link detection is heavily dependent on the timeouts that are used for the links. If it is set too low, the links are deleted before they may be used, even though they might still exist, resulting in a local broadcast on every hop. If it is set too high, links are assumed to exist, but have broken a long time ago.

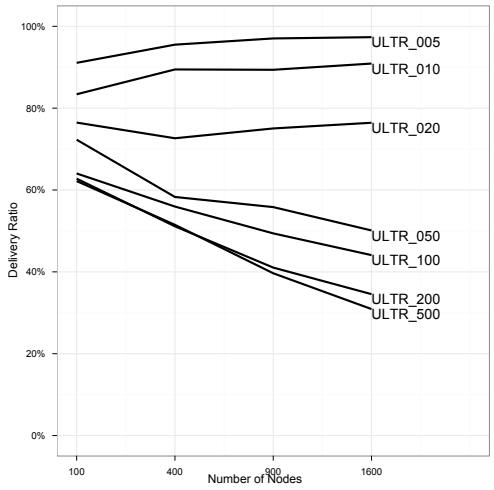


Fig. 9: Delivery Ratio achieved in the Simulations for different timeouts

The implementation of ULTR uses a timer that fires every 100 ms, and has a parameter called `linkTimeout` that defines how many times that timer must fire before a link is removed from the neighbor table. The results presented above were achieved with a `linkTimeout` of 5, and resulted in a lot of message transmissions but also fairly high delivery ratio. To quantify the impact of the `linkTimeout`, the performance of ULTR was measured with different values of `linkTimeout`: 5,10,20,50,100,200 and 500.

Figure 9 shows the delivery ratio achieved by ULTR with the seven different timeouts. It seems that the delivery ratio

 ISBN: 978-1-61208-296-7

is constantly decreasing with increasing timeout lengths. This is not surprising, since a link that has been removed from the neighbor table results in a local broadcast. All nodes that receive this message and know the intended next-but-one hop retransmit the message, adding a lot of redundancy. Therefore, removing a link too early does not result in message loss, but in unnecessary network load. However, if the link is deleted too late, i.e., a link is assumed to exist where it has already broken, the message gets lost. Therefore, when considering only the delivery ratio, using a small timeout seems favorable.

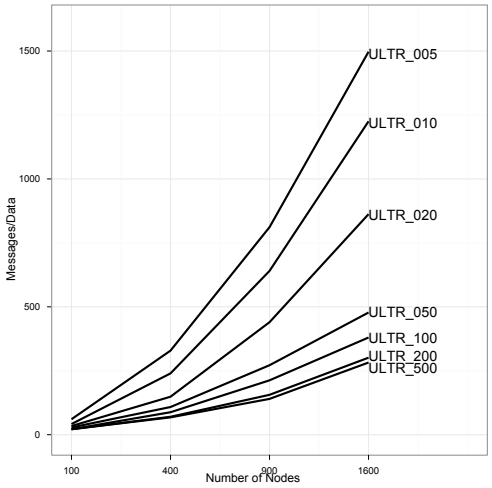


Fig. 10: Number of messages transmitted to deliver a single application message in the Simulations for different timeouts

However, when the network load is considered, the choice seems to be quite the opposite. Figure 10 shows the cost of delivering a single application message measured in transmitted messages. When using the smallest timeout of 5, about 1500 messages are transmitted for each application message delivered in the network consisting of 1600 nodes, which is quite close to the cost of flooding the message. Therefore, the decision which timeout should be used is a tradeoff between delivery ratio and network load. However, there are limits to the choice: Increasing the timeout above 200 does not change delivery ratio or efficiency much. Also, as the delivery ratio is most often more important than the network load, it is unlikely that a timeout of more than 50 would be used, because higher timeout values lead to a delivery ratio of less than 50%. Still, even this is much more than what the related work protocols achieved, making ULTR a fine choice for the evaluated network types.

IV. Related Work

LRS, a link relay service, is proposed in [15]. The authors require all nodes to transmit messages containing their ID and a sequence number regularly, to detect incoming neighbors. Nodes that receive those messages answer with a message of their own, enabling the detection of bidirectional links. For unidirectional or asymmetric links, LRS is used, which floods the messages over a specified number of hops. The main difference between LRS and ULTR is that ULTR uses a chosen forwarding node instead of flooding over multiple hops to circumvent unidirectional links. In the second mode, when neighborhood information is only gathered passively, ULTR does not need to detect neighbors actively as LRS does. Moreover, when a unidirectional link needs to be passed, only nodes that have an active link to the intended next hop forward the message, which once more reduces the network load compared to LRS.

The authors of DEAL [16] describe mechanisms which may be used to detect and exploit asymmetric links in dense wireless sensor networks. They introduce *Source Specific Relay* (SSR), which is used to select a neighboring node X as forwarder, when a link between two nodes A and B is asymmetric. An enhancement, called *Dynamic Driven Maintenance* (DDM) is also described. DDM uses a combination of SSR and broadcast mechanisms to react to changes in the nature of asymmetric links. The third contribution of [16] is called *Asymmetry-Aware Caching* (AAC) and deals with memory requirements on sensor nodes. Due to the limited memory, neighbor tables need to be restrained to a fixed size, leading to eviction of nodes if more neighbors than neighbor table entries are available. Choosing the right entry to evict is far from simple and AAC is used to make this decision, but the exact mechanism is not specified. DEAL is concerned with the detection of asymmetric links on the link layer, ULTR makes use of these links on the routing layer. Moreover, ULTR can also use unidirectional links, which DEAL cannot detect. Finally, ULTR with passive detection can even operate completely without neighborhood management, removing the costs of link detection.

Try-Ancestors-Before-Spreading (TABS) [17] uses any nodes that are closer to the root of a routing tree as forwarders, when the direct parent of a node did not forward the message. To realize this, a node stores a message it has forwarded until the retransmission by its parent node is overheard or a timeout expires. If the timeout expires the message is retransmitted, allowing all nodes that are closer to the sink by a so called *minimum progress limit* to forward the message, regardless of their parent-child relation. The *minimum progress limit* is lowered with each consecutive retransmission, finally reaching a value of zero, indicating that even siblings of the transmitting node may forward the message. Even though this approach increases delivery ratio, it is still necessary for nodes to store messages and wait for acknowledgments. These acknowledgments may only be received if the links are bidirectional, resulting in unnecessary retransmissions if the links are unidirectional. ULTR can use unidirectional links directly, if a neighborhood protocol is provided. If no neighborhood protocol is provided and ULTR uses only passive detection, the forwarding mechanism enables the implicit usage of unidirectional links. Also, no explicit acknowledgments are used in ULTR, reducing the cost further. Moreover, the implicit usage of unidirectional links enables ULTR to work in environments with low link stability, whereas TABS needs fairly stable links for the routing tree. This is also reflected in the used testbed: The changing power supply of battery powered nodes lead to frequent link changes in our testbed, whereas the USB powered testbed used for TABS had a constant power supply, resulting in much more stable links.

ABVCap_Uni [18] uses virtual coordinates to enable geographic routing in sensor networks without geographic information. The authors claim that *ABVCap_Uni* enables the

 ISBN: 978-1-61208-296-7

usage of unidirectional links through definitions of clusters and rings. The overhead of building these clusters and rings is high, though. The simulation used for evaluation did not feature any message losses, and all links were static, leading to a delivery ratio of about 69 - 87%. If the network load induced by constant rebuilding of the clusters and rings would have been included, the delivery ratio would decrease further. As shown in Section III, ULTR was evaluated both in simulations and on a real sensor node testbed.

ieARQ and *E-ieARQ* are introduced in [19]. The authors are concerned with acknowledgment losses due to asymmetric links, which in turn lead to unnecessary retransmissions and a waste of energy. They show that implicit ARQ can lead to an avalanche effect and propose two enhancements of ARQ, which remove the avalanche effect and reduce power consumption by adding an explicit acknowledgment if the implicit one was lost. This approach does not take unidirectional links into account, though, in which case a direct transmission of an acknowledgment is impossible. ULTR operates without acknowledgments or retransmission, removing the avalanche effect completely.

V. Conclusion

Unidirectional links represent a huge problem for routing protocols in wireless networks and especially in sensor networks. Even though experiments show that unidirectional links are quite common, most of today's routing protocols are not able to make use of them. In this paper we introduced ULTR, a routing protocol for wireless sensor networks with often changing and unidirectional links. We presented two versions of ULTR, one which depends on a neighborhood discovery protocol to supply link information and one that uses passive link detection. The former should be used when a neighborhood discovery protocol is included on the sensor nodes anyway. This could be the case when a TDMA MAC is used, which needs to know its two hop neighborhood, when the application needs connectivity data or when a monitoring software is included, which also monitors links and link changes. The latter version, the one that uses passive link detection, should be used when no other source of connectivity data is available.

We evaluated ULTR both in simulations and in real world experiments with eZ430-Chronos sensor nodes from Texas Instruments and compared the performance of ULTR to that of Tree Routing with retransmissions, DSR and AODVBR. The results show that ULTR is much better suited to an environment with often changing links and unidirectional links than the protocols chosen for comparison.

References

[1] S. Lohs, R. Karnapke, and J. Nolte, "Link stability in a wireless sensor network - an experimental study," in *3rd International Conference on Sensor Systems and Software*, 2012.

[2] L. Sang, A. Arora, and H. Zhang, "On exploiting asymmetric wireless links via one-way estimation," in *MobiHoc '07: Proceedings of the 8th ACM international symposium on Mobile ad hoc networking and computing*. New York, NY, USA: ACM Press, 2007, pp. 11–21.

[3] Turau, Renner, and Venzke, "The heathland experiment: Results and experiences," in *Proceedings of the REALWSN'05 Workshop on Real-World Wireless Sensor Networks.*, Jun 2005. [Online]. Available: citeseer.ist.psu.edu/732032.html

[4] J. Zhao and R. Govindan, "Understanding packet delivery performance in dense wireless sensor networks," in *SenSys '03: Proceedings of the 1st international conference on Embedded networked sensor systems*. New York, NY, USA: ACM Press, 2003, pp. 1–13.

[5] M. K. Marina and S. R. Das, "Routing performance in the presence of unidirectional links in multihop wireless networks," in *Proceedings of the 3rd ACM international symposium on Mobile ad hoc networking & computing*, ser. MobiHoc '02. New York, NY, USA: ACM, 2002, pp. 12–23. [Online]. Available: http://doi.acm.org/10.1145/513800.513803

[6] C. E. Perkins and E. M. Royer, ""ad hoc on-demand distance vector routing."," in *Proceedings of the 2nd IEEE Workshop on Mobile Computing Systems and Applications, New Orleans, LA*, FEB 1999, pp. 90–100.

[7] E. Duros, W. Dabbous, H. Izumiyama, N. Fujii, and Y. Zhang, "A link-layer tunneling mechanism for unidirectional links, http://www.faqs.org/rfcs/rfc3077.html," Mar 2001.

[8] S. Mank, R. Karnapke, and J. Nolte, "Mac protocols for wireless sensor networks: Tackling the problem of unidirectional links," in *International Journal on Advances in Networks and Services, vol 2 no 4*, 2009, pp. 218 – 229.

[9] K. Langendoen, A. Baggio, and O. Visser, "Murphy loves potatoes: Experiences from a pilot sensor network deployment in precision agriculture," in *Proc. 14th Intl. Workshop on Parallel and Distributed Real-Time Systems (WPDRTS)*, Apr. 2006.

[10] S.-J. Lee and M. Gerla, "AODV-BR: Backup routing in ad hoc networks," in *Proceedings of the IEEE Wireless Communications and Networking Conference (WCNC 2000)*, Chicago, IL, Sepember 2000. [Online]. Available: citeseer.ist.psu.edu/lee00aodvbr.html

[11] D. Johnson, H. Yu, and D. Maltz, "The dynamic source routing protocol (dsr) for mobile ad hoc networks for ipv4, https://tools.ietf.org/html/rfc4728." [Online]. Available: https://tools.ietf.org/html/rfc4728

[12] A. Varga, "The omnet++ discrete event simulation system," in *Proceedings of the European Simulation Multiconference (ESM'2001)*, Prague, Czech Republic, Jun. 2001.

[13] A. Koepke, M. Swigulski, K. Wessel, D. Willkomm, P. Klein Haneveld, T. Parker, O. Visser, H. Lichte, and S. Valentin, "Simulating wireless and mobile networks in OMNeT++: The MiXiM vision," in *1st Int. Workshop on OMNeT++*, mar 2008. [Online]. Available: http://www.st.ewi.tudelft.nl/ koen/papers/mixim.pdf

[14] "Texas instruments ez430-chronos, http://focus.ti.com/docs/toolsw/folders/print/ez430-chronos.html?DCMP=Chronos&HQS=Other+OT+chronos." [Online]. Available: http://focus.ti.com/docs/toolsw/folders/print/ez430-chronos.html?DCMP=Chronos&HQS=Other+OT+chronos

[15] J. Du, W. Shi, and K. Sha, "Asymmetry-aware link quality services in wireless sensor networks," in *Proceedings of the 2005 international conference on Embedded and Ubiquitous Computing*, ser. EUC'05. Berlin, Heidelberg: Springer-Verlag, 2005, pp. 745–754. [Online]. Available: http://dx.doi.org/10.1007/11596356_74

[16] B. B. Chen, S. Hao, M. Zhang, M. C. Chan, and A. L. Ananda, "Deal: discover and exploit asymmetric links in dense wireless sensor networks," in *Proceedings of the 6th Annual IEEE communications society conference on Sensor, Mesh and Ad Hoc Communications and Networks*, ser. SECON'09. Piscataway, NJ, USA: IEEE Press, 2009, pp. 297–305. [Online]. Available: http://dl.acm.org/citation.cfm?id=1687299.1687333

[17] J. Faruque and A. Helmy, "Tabs: Link loss tolerant data routing protocol for multi-hop wireless sensor networks," *Sensor Networks, Ubiquitous, and Trustworthy Computing, International Conference on*, vol. 0, pp. 11–18, 2010.

[18] C.-H. Lin, B.-H. Liu, H.-Y. Yang, C.-Y. Kao, and M.-J. Tsai, "Virtual-coordinate-based delivery-guaranteed routing protocol in wireless sensor networks with unidirectional links," in *INFOCOM 2008. 27th IEEE International Conference on Computer Communications, Joint Conference of the IEEE Computer and Communications Societies, 13-18 April 2008, Phoenix, AZ, USA*. IEEE, 2008, pp. 351–355.

[19] R. P. Liu, Z. Rosberg, I. B. Collings, C. Wilson, A. Y. Dong, and S. Jha, "Overcoming radio link asymmetry in wireless sensor networks," in *PIMRC'08*, 2008, pp. 1–5.

 ISBN: 978-1-61208-296-7

Combined Time Synchronization and Efficient Data Gathering for Wireless Sensor Networks

Application to Micaz® motes

Jérôme Mathieu
IUT Rodez
University Toulouse 1
Rodez, France
jerome.mathieu@iut-rodez.fr

Vincent Boudet, Jérôme Palaysi
LIRMM
University Montpellier 2
Montpellier, France
vincent.boudet@lirmm.fr
jerome.palaysi@univ-montp2.fr

Sylvain Durand
LIRMM
University Montpellier 3
Montpellier, France
sylvain.durand@lirmm.fr

***Abstract*—This paper presents an algorithm for wireless sensor networks, which combines time synchronization and efficient data gathering. The synchronization part is novel, using multiple ways between the reference and the node to be synchronized to. The data gathering part resumes a previous work. It uses various connected dominating sets for increasing the lifetime. This algorithm is calibrated to Micaz® motes by performing some experiments with them. The synchronization part is then validated on a simple Micaz® network. That shows that our algorithm can effectively merge time synchronization and data gathering with no special message and minimal overhead for synchronization.**

Keywords-Synchronization; Efficient data gathering; Wireless sensor networks; Micaz

I. Introduction

The main specificity of Wireless Sensor Networks (WSN) is due to the limited energy and capabilities (RAM/ROM, processor, etc.) of the motes (nodes). Those limited capabilities require to adopt simple algorithms, especially for annex functions like data gathering. The limited energy requires to search low consumption solutions. One of the best strategies is to reduce the duration of mote's awake periods and the number of communications.

Especially in this constrained context, data gathering needs time synchronization between motes. Otherwise, whether the gathering fails or energy is unnecessarily lost.

Data gathering needs also to route data to the sink point. Therefore, all nodes does not have the same role. Some of them are in the backbone, and then have more jobs to do and de facto consume more energy. The network dies because the backbone nodes deplete their energy if nothing is done.

Many works (see the state of the art in Sections II and III) treat one of the mentioned problems: either efficient routing or synchronization. But, to our knowledge, none of them treats both in a single solution. Consequently, in practice, nodes accumulate communications and jobs to do.

In this article, we propose an algorithm, which combines an existing efficient data gathering [1] and a novel synchronization method, which does not increase the mote awake duration nor the number of messages in comparison to the data gathering performed alone, while minimizing the overhead in messages.

In the following section, we remind the gathering part of the algorithm. In Section III, we explain our novel time synchronization protocol. Section IV presents the specificities of the Micaz® sensors used in the experiments. We present in Section V the algorithm combining gathering and synchronization and its performances in Section VI. This article finishes by giving the conclusions and some perspectives to this work.

II. Efficient Data Gathering

One typical way is to use Connected Dominating Sets (CDS), also called backbones, to route data to the sink. The nodes belonging to a backbone use more energy to forward data because they are much more messages to wait and treat: at least one by child. The goal is then to minimize the number of nodes in a backbone [2]. Because of robustness, scalability and mainly efficiency requirements, most algorithms are operating in a distributed manner with local election of the nodes belonging to the backbone [3]. But, after a certain period of time, the network will be disconnected while the leaf nodes may still have a lot of energy. An improvement (in order to increase the lifetime) is to compute several disjoint backbones [4]. One then tries to use those backbones alternatively. Unfortunately, computing the maximum number of disjoint CDS (called connected domatic number) is a hard task. Furthermore, this number may be very small. For example, Islam et al. [5] show that for a grid graph where one of the dimensions is greater or equal to 3, the connected domatic number is 1. Thus in such a graph, one cannot expect to increase the lifetime of the network using disjoint CDS.

The disjoint constraint may be relaxed by trying to find a set of CDS such that the maximum number of CDS a node belongs to is minimized. Such a distributed algorithm is proposed in [6]. Nevertheless, this model does not take into account the real consumption of energy of the nodes, which depends (among others) on the number of received messages, i.e., on the degree of the node.

In our case, all data have to be gathered to a fixed sink. We thus only have to compute a directed in-tree rooted at the

 ISBN: 978-1-61208-296-7

sink, and the sink may initiate this computation (the algorithm is not localized but only distributed). This specification allows us to need only a small number of messages to compute a backbone.

III. Time Synchronization

In a perfect data gathering, all the nodes receive data from their children simultaneously and send data to their parent a short time later. After that, nodes can enter in a sleeping period until the next gathering. Without any time synchronization, all nodes must wait each other, canceling the sleeping period. Of course, the greater the sleeping period duration is, the longer the lifetime of the network is. Unfortunately, accurate time synchronization protocols [7] imply exchange of many messages between nodes and a lot of calculations, which reduce the sleeping period duration. Furthermore, nodes cannot send messages simultaneously because of wireless media access control protocols. So, in our point of view, a good time synchronization protocol does not need to be very accurate but should not add messages and should not reduce the sleeping period duration.

Typically, each node must evaluate two parameters [8]: the offset, i.e., the difference between the reference clock and the node clock, and the skew due to the drift of a node clock relative to a perfect clock. Offset is the leading parameter in the short-term, for example because nodes are not switched on simultaneously. Skew becomes important in the long-term, for example when data gatherings are infrequent.

A. Offset computation process

Offset is calculated [8] whether by a round-trip between the reference node and an evaluated node, or after the estimation of the delay D involved in the transport of a message between these two nodes. The first type of protocols supposes that D is equal in each way. The accuracy of all methods depends on the quality measurements of D. In practice, D varies from one message to the other and is in part unpredictable. That's why all methods use statistics to improve the offset estimation. But, except for RBS [9], these statistics are based on repetitive exchanges of messages between always the same nodes. The result is assigned to a specific node. The complexity increases when the evaluated node is not directly reachable by the time reference node. Classically [7,9], the synchronization is then achieved in stages, some nodes being elected as intermediate time reference.

Our strategy, inspired by RBS, is different from two points. First, if the network is homogeneous – i.e., all nodes have the same architecture, execute the same program, etc. - statistics are calculated spatially instead of temporally. Second, the evaluated node doesn't need to be directly reached by the time reference node. Suppose a simple network as shown in Figure 1. Node 0 sends by broadcast a message M_1 at time Tr, so M_1 is received approximately at the same time by node 1 and by node 2 (The difference in distance between nodes involves a negligible time difference). Node 2 sends the message M_2 immediately after receiving M_1. Node 1 receives M_1 at $T[0]$ and, a few later, M_2 at $T[1]$. Then, from the point of view of node 1, $D=T[1]-T[0]$. In order to do so, D must be considered as a constant in the network.

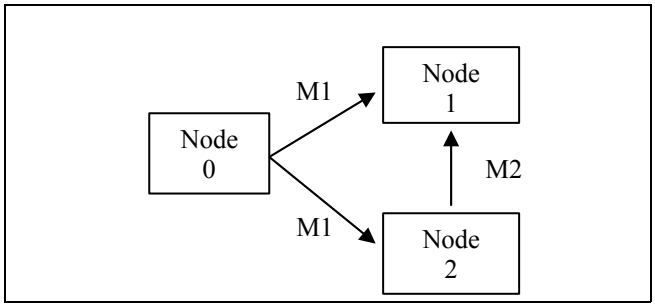


Figure 1. Synchronization by two ways

If fact, D is the difference of the delay between nodes 0 and 1 passing by node 2 (=2D) and the delay between nodes 0 and 1 (=D). If the time reference node is node 0, the offset of node 1 is $T[1]-T[0]-D$. So, the reference time Tr must be incorporated at least in M_1.

Our method requires to calculate a tree, which is necessary for data gatherings anyway, and each node must have at least two neighbors. The synchronization can be done along with the tree calculation. The reference time is that of the root of the tree, i.e., the sink.

Formally, the evaluated node receives a message from its parent (which is at a depth $d[0]$ in the tree) at time $T[0]$ and from other neighbors (depth $d[i]$) at time $T[i]$. All $d[i]$ must be different from $d[0]$. Then, the delay D from the point of view of the evaluated node is:

$$D = \text{mean}((T[i]-T[0]) / (d[i]-d[0])) \quad (1)$$

and its offset can be calculated by:

$$\text{offset} = \text{mean}(T[i]-Tr-d[i]*D) \quad (2)$$

This is done each backbone calculation.

B. Skew computation process

The skew is often supposed to be a constant [7,8,9], at least over a period of a few minutes. Typically, something like the offset variations over time are fitted by a line (linear regression, for example). The skew is the slope of this line. For that, many offsets are calculated in a relatively short time. Then, a good skew estimation requires a very high offset accuracy and especially much more messages consuming the battery.

We propose to evaluate the skew using data gatherings messages only. In each data gathering, a node sends its data to its parent node in the tree. To ensure the sending, the parent node returns an acknowledgment to its child node. This <ACK> message contains the time of the parent node. Then, the child node can follow over time its clock drift regarding its parent node without any additional message. Moreover, the skew value is not important in itself. The main

 ISBN: 978-1-61208-296-7

objective of a node is to keep synchronized to its parent over time. For that, if $\Delta T[i]$ is the difference between a child's time and its parent time observed at the i^{th} gathering, then $\Delta T[i+1]$-$\Delta T[i]$ is more important (equal to the skew times the time interval between two gatherings). For the first gatherings, our algorithm provides sufficient margin of errors from experimental results.

IV. The Micaz® motes

The objective is to implement our algorithm in the Micaz®'s platform. Some steps of it are adjustable, depending on the platform.

A. Energy consumption

The transceiver (radio model TI CC2420) is the component that has the highest energy consumption of all relevant components of the Micaz® [10]. In particular, the receive mode consumes 19.7 mA while the processor needs "only" 8 mA in its active mode. When the transceiver is in the transmit mode, the current consumption varies from 8.5 to 17.4 mA depending on the chosen transmit power. Then, the total consumption is always about 28 mA. So, for energy saving, the transceiver must be off as often as possible. That means reducing the number of messages and the time it may be used.

The reference objective of this study is one gathering every minute and a lifetime of one year (a directive of a partner). A node in the economical state consumes approximately 20µA (CC2420 power down mode) + 15µA (CPU save mode). The consumption of other components is neglected here. In one year, an idly node consumes 35µA times 8760h (1 year), that is about 307 mAh. A Micaz® embeds two AA batteries. If their capacity is 2200 mAh each, the lifetime in active mode is about (4400-307)/28 = 146 h. If this lifetime is splitted, each data gathering (with or without backbone calculation) must take at most 1 s.

Note that the CC2420 crystal oscillator start-up time is 1 ms, the condition to switch from the power down mode to the idle mode from which the communications can restart. But the idle mode consumes too much (about 0.43 mA). It is an another proof that energy saving is not completely compatible with high accuracy synchronization protocols.

To validate these calculations, one could use an accurate tool of energy consumption prediction [11]. But, to be really useful here, the tool should also integrate a model of battery. For example, Alkaline batteries, such they use, have a cut-off voltage smaller than the minimum voltage supply for Micaz® and their capacity decreases if the load increases [12]. All of that make the battery lifetime prediction hazardous. So, the mote lifetime is evaluated by some experiments. A Micaz® is programmed to send a message every minute and to wait the rest of the time. The transceiver is switched off a percentage of the duty cycle, after the sending. A LED is blinking 5% of the duty cycle, only for control. It represents an additional amount of consumed energy of about 0.5%.

The lifetime of the mote is reported in Table I. We can see that the lifetime of the mote fully active is not a constant because the behavior of the alkaline batteries. But, its tendency suggests that 146 h is probably easily reachable especially when the active time of the transceiver will be less than 1 %.

TABLE I. Evolution of the Mote Lifetime vs the Percentage of the Active Time of the Transceiver

Active time of the transceiver (%)	100	75
Mote lifetime (h)	113.80±0.66	178.09±1.52
Mote fully active lifetime (h)	113.8	133.6

B. Communication delays

Firstly, our algorithm has a flaw: when a node has only one neighbor (its parent), the delay D can't be calculated as shown above. It must be given as a platform constant for the offset calculation process. But the skew correction process finely corrects the local clock, if necessary, at the first data gathering.

Secondly, in the data gathering phases, backbone's nodes have to wait all their child nodes before sending its data to its parent node. A timer (timeout) is necessary otherwise a deadlock situation may occur if one child node fails. The timer setting is difficult both to ensure the time needed to gather data and to save energy.

The following experiments are used to evaluate the real delays D depending on the number of neighbors. A Micaz® node, randomly chosen, called the evaluated node *eN*, is programmed to send a broadcast message that contains its system time S_s. Neighbor(s) send(s) back this message without any modification (echo) to *eN* as soon as possible. The nominative sendings allow not saturating unnecessarily the neighbors. At the reception of the echo by the neighbor *Vi* (one per neighbor), *eN* gets as soon as possible its associated system time $S_e[Vi]$ and sends it back to *Vi* in a <ACK> message. Neighbors have a timer associated with the sending of the echo. If the timer fires without receiving this <ACK>, they send echo again. Each neighbor receiving this <ACK> message stops until the next test. This ensures that *eN* has processed all neighbors. It is necessary because neither CSMA/CA nor the beacon mode are incorporated by default in Micaz®. We only have the clear channel assessment. At the end of a trial, *eN* sends S_s and all $S_e[Vi]$ to a wireless network "sniffer". This "sniffer" captures all exchanges too. $S_e[Vi]$-S_s is the *eN*'s delay D plus the *Vi*'s one.

This procedure is repeated 30 times, every 5s, for statistical analysis. The repetition rate is chosen to minimize the influence of the clock drift while leaving time for nodes to prepare for the next round.

The system time accuracy is estimated at 0,25 ms.

Table II shows that the nodes, even placed in the same situation, do not necessarily behave in the same way (see 1 neighbor column). When the number of neighbors increases, the mean delay, the standard deviation and the max value increase but not evenly for all nodes. Nodes little bit more reactive maintain relatively low values, but others not. With 3 or 4 neighbors, it is better to forecast 100 ms to gather data.

 ISBN: 978-1-61208-296-7

Otherwise, one of the nodes does not have the time to send data to its parent. This table gives also an idea of the needed synchronization accuracy. It seems that accuracy around 5 ms is enough.

We must specify that all these results depend on the chosen timer. Indeed, it is one of our future works to refine the timer formulation.

TABLE II. EVOLUTION OF DELAYS VS THE NUMBER OF NEIGHBORS

Number of neighbors	Evaluated node's delay + neighbor's delay			
	Neighbor	*Mean (ms)*	*Standard deviation (ms)*	*Max value (ms)*
1	"1"	3.1	2.1	7.3
	or "2"	5.0	3.2	13.0
2	"1"	6.6	9.3	39.1
	"2"	5.4	8.2	41.6
3	"1"	11.5	20.2	104.9
	"2"	8.0	9.0	30.3
	"3"	11.3	17.7	89.2
4	"1"	6.3	7.2	30.1
	"2"	9.4	9.6	38.7
	"3"	7.6	7.6	29.6
	"4"	12.0	20.6	97.8

Table III shows the most important consequence when the number of neighbors increases. Many messages are lost by *eN* because it is busy by treating other message just received.

TABLE III. EVOLUTION OF PERCENTAGE OF LOST MESSAGES AND RETRANSMISSIONS VS THE NUMBER OF NEIGHBORS

Number of neighbors	Percentage of lost messages	Percentage of retransmissions
1	0	0
2	10	16 to 20
3	20	26 to 38
4	32	40 to 70

C. Clock skew

Two motes are programmed to send a message every minute to the "sniffer". Then, the mote's clock can be compared to the computer's clock, which is supposed to be accurate on the experiment's duration (about 80 h). The first mote skews of +6.6 s (about +2.0 s per day or +23 ppm), while the second one skews of +4.6 s (about +1.4 s per day or +16 ppm). These results corroborate some previous works [13].

V. THE PROPOSED ALGORITHM

Our algorithm includes two phases:

- a backbone calculation + synchronization phase,
- a data gathering + skew correction phase.

The first phase is repeated using several conditions. The basic condition is the number of gatherings: each *X* gatherings, a new backbone calculation is initiated by the sink node. Other conditions to initiate a new backbone calculation are relative to data loss (just one data or beyond a percentage threshold). The choice of these conditions depends on the application. The choice of *X* (not necessary a constant) depends on the percentage of energy consumed during a gathering phase.

The second phase must incorporate a signaling procedure if a condition on data loss is used. Indeed, leaf nodes in particular do not know if a gathering is complete or not, and consequently if a new backbone calculation will be initiated or not. Between 2 first phases, gatherings (second phases) are repeated with a rhythm depending on the application.

A. The backbone calculation + synchronization phase

This phase also incorporates the neighbor discovery. This is done by observing the ID of each received message.

There are 3 kinds of messages:

- <INV>: the "invitation" message, broadcasted. It contains the backbone ID *B*, the sender ID, the sink time *Ts*, the total waiting in the path *W*, the sender level in the tree *L*, etc.
- <F>: the "parent" message, sent repeatedly to the parent node until <ACK> is received, contains *B*, the sender ID, etc.
- <ACK>: the acknowledge message, not mentioned in Figure 2.

The principle of the backbone calculation is that each node will choose as a parent in the backbone the first node from which it received an "invitation" message. Since all the nodes but the sink send their invitation after receiving one, this ensures that we built a directed in-tree rooted at the sink. Fine-tuning of the algorithm is done by the computation of the delay *w* each sensor has to wait before sending an invitation. The main idea is that less remaining energy the node has, longer the delay is. Of course, if the delay increases, the probability for a node that its <INV> will be the first received by its neighbors decreases. Thereby, the probability that the node belongs to the backbone decreases too. This tends to calculate different backbones and equalize the energy consumption of nodes. The main consequence is an increasing of the network lifetime. More details can be found in [1].

Note that if the delay is constant for all nodes, then this phase is formally identical to a Breadth First Search (BFS) with incorporated synchronization.

Finally, it is noteworthy that no additional message is required to ensure the synchronization. The three ones are necessary for the backbone calculation. And the additional computations are very light.

B. The gathering + skew correction phase

The backbone nodes wake up at the same time as its child nodes because they must be ready to receive <DATA>. This time is depending on the level in the backbone,

 ISBN: 978-1-61208-296-7

following the staggered sleep scheduling scheme like in D-MAC [14] but at the application level. Working at the application level instead of the MAC level allows to reduce the transmission slot duration as soon as possible.

For the sink node (also the reference time node) *S*, send <INV> ; *B*++

For all other nodes *N* do:

- **Receive the first <INV>**:
  (parent ID = sender's ID) and (*N*'s *L* = *L*+1)
  Note *Ts*, *L*[0]=*L*, *T*[0]=*T*-*W*, where *T* is the local system time
  Send <F>
  Chose a delay *w*; *W* = *W* + *w*; wait *w*
  Send <INV>

- **Receive other <INV>**:
  If (*L* ≠ *L*[0]), *L*[*i*]=*L*, *T*[*i*]=*T*-*W*, *nb_inv*++, *i*++

- **Receive a <F>**:
  N ∈ backbone

- **Just before go sleeping**:
  If (*N* is a parent of none node) then *N* ∉ backbone
  If (*nb_inv* > 0) apply Equation (1) else *D* = 5 ms
  Apply Equation (2) (with *Tr* = *Ts*)

Figure 2. The backbone calculation + synchronization phase

For the leaf nodes do:
Send <DATA> to its parent repeatedly until <ACK> is received
T = *Tf*
If (ALERT=Yes) for the first time, return to the backbone calculation + synchronization phase in *C* gatherings
else go sleeping until the next gathering

For the backbone's nodes do:
Receive <DATA> from a child *N*
Send <ACK> with its local system time *Tf* and ALERT (ALERT=No by default) to *N*
Wait for the end of the gathering period (timer)
Behave as a leaf node

Figure 3. The gathering + skew correction phase

The simplest and energy saving manner for signaling a new backbone calculation is to use an ALERT flag (see Figure 2). ALERT is incorporated in <ACK> messages, associated to the maximum depth in the tree (*DT*) and the sender's time, i.e., the parent's time *Tf*.

It allows to change the node phase together with other nodes through the counter *C*=*DT*-*L*. Note that *DT* gatherings are necessary before effectively initializing a new backbone calculation. It is the main inconvenient of this method. Another solution, when data are vital, is to keep all nodes pending the decision of the sink. That means that all nodes wait the end of the gathering for the ALERT signal, consuming uselessly their energy.

The affectation *T*=*Tf* may seem a bad formula to correct the skew. Normally, the delay *D* must be taken into account. But *D* may overestimate the real transmission time because the context is usually more favorable. Then, at the next gathering, the child may wake up before its parent, failing the firsts sendings of data. It is thus preferable that leaf nodes wake up after their parent.

Once again, no additional message and few additional computations are necessary for the synchronization. The main part of the work is needed to ensure the dependability, the reliability of the data gathering.

The gathering period should last the time required for all children to send their data, that is for example about 100 ms for a network with 4 neighbors per node (see table II). Then, a node in the backbone awakes about 115 ms (mean time) while a leaf node awakes only about 15 ms (mean time) (see table II). That is the reason for changing regularly the backbone.

A time of 115 ms fulfills the recommendations made in Section IV-A. If more than one gathering per minute is expected, the gathering period should be decreased, i.e., the medium access control should be improved (see Section IV-B).

VI. Validation of the Synchronization Protocol on Micaz®

The backbone calculation + synchronization (BCS) phase (Figure 2) is programmed on 5 motes. Another node is programmed as a sink node. The nodes are switched on randomly. The BCS phase starts automatically about 30 s later the sink node has been switched on. For the experiments purposes, each node has an additional functionality: responding to a probe node. When the BCS phase is ended, the probe node sends (broadcast) a special message to all nodes including the sink node. Immediately, nodes have to record their local time. Then, they have to send it to the "sniffer". Clocks, corrected by the calculated offset, are compared to the sink's one, which is the time reference.

The experimental network is presented Figure 4. It is constructed such that several situations are treated:

- A node has only one neighbor (node 5), so *nb_inv* = 0 and *D* is fixed to 5 ms for this node (see Figure 2).
- Some nodes have 2 neighbors, at its same level or a smaller level (nodes 2 and 4).
- Some nodes have many - 3 or 4 - neighbors (nodes 1 and 3).
- The backbone has several levels (3), which is the maximum achievable with this experiment and only 2 "sniffers".
- The backbone is always the same to make statistics without influence of the resulting backbone.

 ISBN: 978-1-61208-296-7

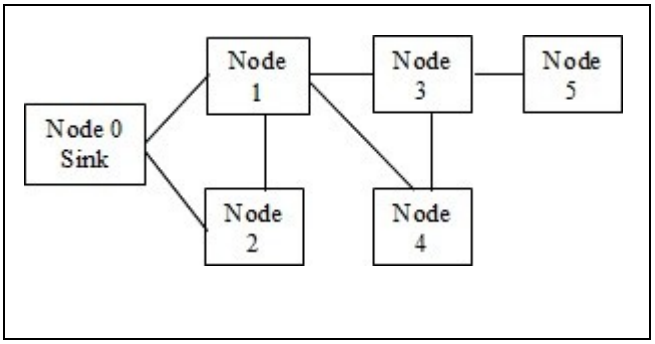


Figure 4. Experimental network

The results are presented in Table IV. They are good, even if they are probably worse than those that have been obtained with other synchronization protocols. They are of the same order of the measured delay shown in Table II, and this is what we hoped, even if the system time accuracy is worse (slower clock rate).

TABLE IV. CLOCK DIFFERENCES BETWEEN THE SINK NODE AND THE OTHER NODES

Node	Mean of clock differences (ms)	Standard deviation (ms)
1	6.6	5.27
2	1.0	8.19
3	3.2	7.33
4	2.4	11.76
5	7.6	3.58

Some remarks should be noted.

First, as expected, and unlike other protocols, the mean of the clock differences does not significantly increase with the depth of the node in the tree. However, our results are almost always positive implying that the delay is probably systematically underestimated. This is probably due to computation time that is not properly taken into account. An optimization of the program should solve the problem.

Second, the number of neighbors influences differently the mean and the standard deviation. When a node has many neighbors, the mean tends to increase and the standard deviation tends to decrease. This shows that our method tends to converge, even if it is towards a value slightly underestimated (cf. the first remark, with cumulative effects).

Third, node 5 is a special case. Its particularly low standard deviation shows that the real delay is fairly stable from one experiment to another. For cons, its mean shows that the default value given to D is also slightly underestimated.

VII. CONCLUSION AND FUTURE WORKS

In this paper, we presented a new protocol that realizes both gathering and a new synchronization method in a wireless sensor network. The load for the gathering is distributed over all the nodes and there is no special message needed for the synchronization.

The perspectives of this work are to refine the timer's formulations and mainly the sleep scheduling for a more deterministic one. Thereafter, the protocol will be tested with a more complex experimental network.

REFERENCES

[1] V. Boudet, S. Durand, L. Gönczy, J. Mathieu, and J. Palaysi, “Efficient gatherings in wireless sensor networks using distributed computation of connected dominating sets”, Sensors & Transducers Journal, vol. 14-2, 2012, pp. 297-307.

[2] S. Guha and S. Khuller, “Approximation algorithms for connected dominating sets”, Algorithmica, vol. 20, April 1998, pp. 374-387.

[3] C. Adjih, P. Jacquet, and L. Viennot, “Computing connected dominated sets with multipoint relays”, Adhoc & Sensor Wireless Networks, vol. 1, no. 1, 2004, pp. 27-39.

[4] T. Moscibroda and R. Wattenhofer, “Maximizing the lifetime of dominating sets”, Proceedings of the 19th IEEE International Parallel and Distributed Processing Symposium (IPDPS'05) - 5th IEEE International Workshop on Algorithms for Wireless, Mobile, Ad Hoc and Sensor Networks (WMAN 2005), Denver, USA, vol. 13, April 2005, p. 242b. doi: 10.1109/IPDPS.2005.276.

[5] K. Islam, S. G. Akl, and H. Meijer, “A constant factor localized algorithm for computing connected dominating sets in wireless sensor networks”, Proceedings of the International Conference on Parallel and Distributed Systems (ICPADS'08), Melboune, Australia, 2008, pp. 559-566.

[6] K. Islam, S. G. Akl, and H. Meijer, “Distributed generation of a family of connected dominating sets in wireless sensor networks”, in Distributed Computing in Sensor Systems, vol. 5516, B. Krishnamachari, S. Suri, W. Heinzelman, and U. Mitra, Ed., Springer Berlin / Heidelberg, 2009, pp. 343-355.

[7] S. el Khediri, N. Nasri, M. Samet, A. Wei, and A. Kachouri, “Analysis study of time synchronization protocols in wireless sensor networks”, Int. J. of Distributed & Parallel Systems, vol. 3, no. 3, May 2012, pp. 155-165.

[8] F. Sivrikaya and B. Yener, “Time synchronization in sensor networks: a survey”, IEEE Network, vol. 18, issue 4, July 2004, pp. 45-50.

[9] J. Elson, L. Girod, and D. Estrin, “Fine-grained network time synchronization using reference broadcasts”, Proceedings of the 5th symposium on Operating systems design and implementation (OSDI'02), Boston, USA, December 2002, pp. 147-163.

[10] http://www.memsic.com/userfiles/files/Datasheets/WSN/micaz_datasheet-t.pdf

[11] O. Landsiedel, K. Wehrle, and S. Götz, “Accurate prediction of power consumption in sensor networks”, Proceedings of the 2nd IEEE Workshop on Embedded Networked Sensors (EmNetSII), Sydney, Australia, 2005, pp. 37-44.

[12] http://en.wikipedia.org/wiki/Alkaline_battery

[13] M.B. Uddin and C. Castelluccia, “Toward clock skew based wireless sensor node services”, Proceedings of the 5th Wireless Internet Conference (WICON) - International Workshop on Ubiquitous Body Sensor Networks (UBSN 2010), Singapore, March 2010, pp. 1-9.

[14] G. Lu, B. Krishnamachari, and C. Raghavendra, “An adaptive energy-efficient and low-latency MAC for data-gathering in sensor networks”, Proceedings of the 18th IEEE International Parallel and Distributed Processing Symposium (IPDPS'04) - 4th IEEE International Workshop on Algorithms for Wireless, Mobile, Ad Hoc and Sensor Networks (WMAN), Santa Fe, USA, April 2004. doi: 10.1109/IPDPS.2004.1303264.

 ISBN: 978-1-61208-296-7

A Scalable Localization Scheme using Particle Swarm Approach for Sensor Networks

Pei-Hsuan Tsai
Institute of Manufacturing Information and Systems
Department of Computer and Information Engineering
National Cheng Kung University
Tainan, Taiwan R.O.C.
Email: phtsai@mail.ncku.edu.tw

Chun-Lung Lin
Industrial Technology
Research Institute
Hsinchu, Taiwan R.O.C.
Email: chunlung@itri.org.tw

Ching-Yi Chen and Jia-Shung Wang
Department of Computer Science
National Tsing Hua Univeristy
Hsinchu, Taiwan R.O.C.
Email: jswang@vc.cs.nthu.edu.tw

***Abstract*—This paper proposes a scalable grid-based localization scheme for establishing relative coordinate system of sensor networks. A distributed selection algorithm is proposed to select a small number of nodes from the set of sensor nodes as virtual grid nodes. The virtual grid nodes are then used to establish grid coordinate system by using particle swarm optimization. The grid coordinate system forms the backbone of node localization. The other nodes, which are not selected as grid nodes, can then compute (distributed) their locations based on the grid coordinate system efficiently. The precision of node localization is highly dependent on the number of grid nodes and can be adaptively adjusted according to target applications. Furthermore, the proposed localization scheme has high scalability and can work without any assistance of GPS. The simulation results show that about sixty percent of non-grid nodes could be correctly positioned in terms of grid coordinates, and nearly all of the non-grid nodes could be correctly be located in one of nine grids surrounding their real positions.**

***Keywords*—wireless sensor networks; localization; grid coordinate system.**

I. INTRODUCTION

In recent years, wireless sensor networks have become an active research topic because of its wide variety of applications, such as battlefield surveillance, smart environments, healthcare, disaster relief [1], [2]. Wireless sensor networks also draw a lot of research challenges, such as sensing coverage controls, energy-efficient communication protocols, and data aggregations. The problem of localization aims to compute the coordinates of nodes in the network. Geographic information can be obtained by installing a Global Positioning System (GPS) receiver on each node in the network. Feasible solutions, however, are to equip only a few nodes with GPS, called *anchor nodes*, and the other nodes, using the estimated distances between anchor nodes and them to infer their position. Extensive studies have shown that the more estimated distances used to compute the coordinates of nodes, the higher accuracy of the established coordinate system, because measurement errors in the estimation of distances can be more effectively eliminated. However, the computational complexity of establishing the coordinate system increases exponentially with the number of distances used in calculating the coordinates of nodes. Although location information is crucial in wireless sensor networks, it is not necessary for all nodes in the network to know their precise position. In fact, requirements of location accuracy is highly application dependent in wireless sensor networks. For example, for many location-based routing protocols [3]–[6], grid-based coordinate system provides sufficient geographic information. Another example is that, in many event-based applications, we are interested in the region where an event takes places rather than the precise location of the node issuing the event. In this case, region-based coordinates are accurate enough for supporting target applications.

In this paper, we propose a scalable grid-based localization scheme for establishing relative coordinate system of sensor networks. The basic idea of the localization scheme is to establish a grid-based coordinate system by using a small number of virtual grid nodes selected from a given set of sensor nodes (In the following, those nodes that are not selected as grid nodes will be referred to as non-grid nodes). The grid coordinate system acts as the backbone of localization for the non-grid nodes. That is, the non-grid nodes can compute their positions via information exchange with their adjacent grid nodes. The overall precision of node localization is highly dependent on the number of grid nodes. The more the number of grid nodes, the higher the accuracy of localization. In such way, the localization scheme can be adaptively adjusted to balance the tradeoff between computation complexity and precision of the established coordinate system according to target applications. In summary, the proposed localization scheme has the following features:

1) **Adaptive**: The precision of the grid-based localization technique can be adjusted adaptively according to the requirements of target applications so that all nodes in the network could be positioned efficiently.
2) **Self-configurable**: The grid-based localization technique establishes the local grid coordinate system without the existence of anchor nodes (the sensor nodes that know their global coordinates). The grid coordinates, however, can also be transformed into the global coordinates if there are at least three grid nodes knowing their global position.
3) **Scalable**: The grid-based localization scheme works in large-scale sensor networks, e.g., more than 500 nodes.

The rest of this paper is organized as follows. Section II introduces the related works. In Section III, the proposed

 ISBN: 978-1-61208-296-7

localization scheme is presented. The simulation results are given in Section IV. Finally, we draw the conclusion of this paper and the future work in Section V.

II. Related Works

DV-Hop approach assumes that the network comprises a small number of anchor nodes [7]. The other non-anchor nodes calculate their locations based on the averaged one-hop distance to every anchor nodes in the network. Leng et al. [14] addressed the problem of sensor localization in wireless networks in a multipath environment and proposed a distributed and cooperative algorithm based on belief propagation, which allows sensors to cooperatively self-localize with respect to a single anchor node in the network, using range and direction of arrival measurements. The research study [8] investigated neighborhood collaboration based distributed cooperative localization of all sensors in a particular network with the convex hull constraint. Three iterative self-positioning algorithms were present for independent implementation at all individual sensors of the considered network. Gu et al. [9] addressed the localization with incompletely paired mutual distances and proposed a Partially Paired Locality Correlation Analysis (PPLCA) algorithm. Sugihara and Gupta [10] described a novel SDP-based formulation for analyzing node localizability and providing a deterministic upper bound of localization error. The SDP-based formulation gives a sufficient condition for unique node localizability for any frameworks. A compressive sensing (CS) based approach for node localization in wireless sensor networks was presented by Zhang et al. [15]. Low et al. [16] proposed a localization system in which localization information is obtained through a probability based algorithm that requires the solving of a nonlinear optimization problem. The authors used the particle swarm approach to solve the nonlinear optimization problem.

III. The Proposed Grid-based Localization Algorithm

Given a set of sensor nodes S deployed in an interested region F, the coordinate system is established in two steps. First, a small number of nodes are selected as grid nodes to establish the grid coordinate system by using the proposed distributed selection algorithm. The grid nodes compute their grid coordinates themselves by exchanging information between each other and minimizing an error objective function. In this paper, a Particle Swarm Optimization technique is proposed to minimize the error objective function and to establish the grid coordinate system. The grid coordinate system then serves as the backbone of the coordinate system, whereby the other nodes in the network can compute their own position according to the grid coordinate system.

A. Grid Coordinate System Establishment

1) Distributed Grid Node Selection Algorithm: To create the grid coordinate system, all the nodes in the network need to estimate the distances between themselves and their communication neighbors. There are several technologies that can be used for this purpose, such as Received Signal Strength (RSS) and Time-of-Arrival (ToA) [17]. Generally, RSS is easier to implement, while ToA may have higher accuracy. Let $\tilde{d}_{i,j}$ be the estimated distance between two adjacent nodes s_i and s_j. Let w be the grid width of the grid coordinate system. We assume, without loss of generality, sensor nodes are deployed in a $L \times L$ square region, where L is multiple of w. To improve computation efficiency as well as localization accuracy, if the estimated distance between two adjacent nodes is smaller than the grid width w, only one of them is selected as a grid node for establishing the grid coordinate system. Thus we need to find the maximum set of nodes so that the distance measurement between any two nodes in the set is larger than or equal to the grid width w. Let $G = (V, E)$ be a graph. Each node $s_i \in S$ is represented by a vertex $v_i \in V$. There is an edge $e \in E$ between two vertices v_i and v_j if and only if $\tilde{d}_{i,j} < w$. An independent set of graph G is a subset $V' \subseteq V$ vertices such that each edge in E is incident on at most one vertex in V'. A maximal independent set is an independent set V' such that for all vertices $v \in V - V'$, the set $V \cup \{v\}$ is not independent. The selection of maximum set of grid nodes for creating the grid coordinate system is equivalent to one of finding the maximum independent set of the graph G, which is a NP-complete problem.

In this paper, we propose a distributed grid node selection algorithm, similar to the one proposed by Luby [18]. Let N_i be the set of nodes that are within the transmission range of node s_i. Let $NG_i = \{v_j \| \forall v_j \in V \text{ and } \tilde{d}_{i,j} < w\}$ be the set of nodes that their distances to s_i is smaller than the grid width. The distributed grid node selection algorithm proceeds in rounds. In each round, every node sends to all its neighbors a message containing its identity. When a node s_i collects all such messages from its neighbors, it becomes a grid node (winner) if either of the following two conditions is satisfied:

1) NG_i is nonempty (there exists at least one communication neighbor s_j such that $\tilde{d}_{i,j} < w$). Moreover, s_i has the smallest Node ID when comparing its Node ID with the received Node IDs of the nodes in NG_i.
2) NG_i is empty

Subsequently, grid nodes (winners) send to all their neighbors a message specifying their winner state. A sensor that receives such a message from one of its neighbors becomes a non-grid node, loser. Winners and losers of a round do not participate in subsequent rounds. The expected number of rounds is $O(logn)$ where n is the number of initial participants. An analysis of the expected number of rounds and a proof of termination of a similar algorithm can be found in Luby [18].

2) Particle Swarm Optimization (PSO) Algorithm: Let S' be the set of grid nodes obtained by using the proposed distributed grid node selection algorithm. In order to establish the grid coordinate system, it is crucial to obtain the distance measurement between two grid nodes in S'. If two grid nodes are adjacent, ToA or RSS can be employed to estimate their distances. However, the distance estimation becomes nontrivial when the two grid nodes are not within the transmission range of each other. In this paper, we employ a simple and effective scheme to estimate the distance between two remote grid nodes, called multi-hop distance estimation model [15]. The model utilizes the correlation of the Euclidean distance and the corresponding shortest path length between two nodes in the network for a given node distribution. Based on this model, a node s_i can estimate the Euclidean distance to another node s_j by sending a control packet that includes a *route length field*

 ISBN: 978-1-61208-296-7

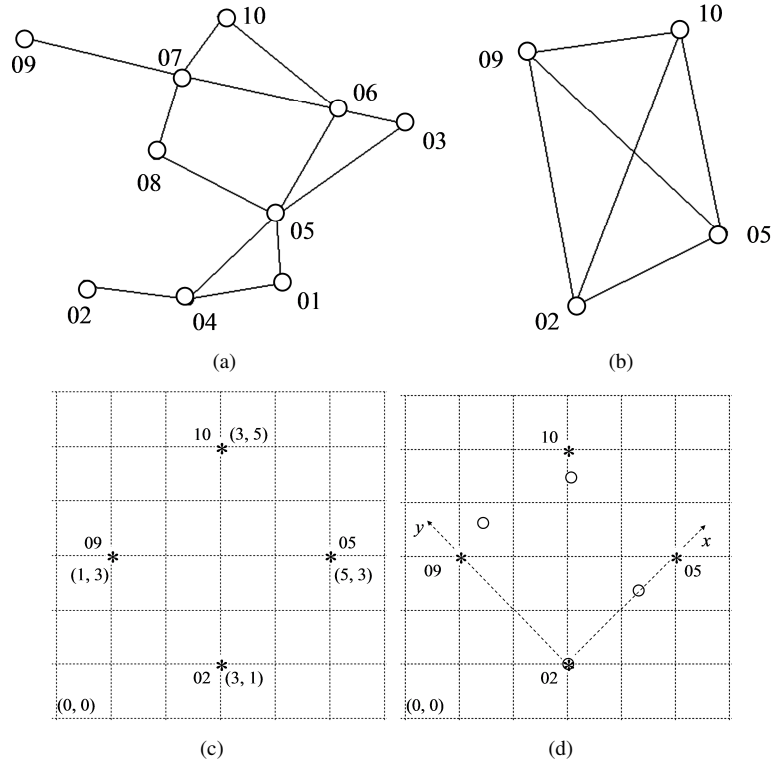


Fig. 1. Grid coordinate system establishment. (a) The constructed graph G(V, E) (b) The set of grid nodes for establishing the grid coordinate system. (c) The local grid coordinate system. (d) The transformed grid coordinate system.

with initial value of zero. When an intermediate node receives the control packet, it adds the one-hop distance between itself and the previous node, obtained by using ToA or RSS, to the route length field. Upon receiving the control packet, node s_j sends it back to node s_i. After node s_i receives the return control packet from s_j, it can read the route length field and estimates the distance between s_i and s_j according to the multi-hop distance estimation model. Here, we also use $\tilde{d}_{i,j}$ to denote the estimated distance, obtained by using the multi-hop estimation model, between two remote grid nodes s_i and s_j. (If two nodes are adjacent, $\tilde{d}_{i,j}$ denotes the estimated distance obtained by using ToA or RSS). We assume, without loss of generality, the grid node with the lowest identify number, say s_1, will perform the calculation for creating the grid coordinate system. Once a grid node s_i has obtained a set K_i of distances to all other grid nodes, i.e., $K_i = \{\tilde{d}_{i,j} \| \forall s_j \in S' \text{ and } j \neq i\}$, it sends K_i to s_1. Thus s_1 has the distance information between any two grid nodes in the network. Our objective is to calculate the grid coordinate system that minimizes the sum of the errors of the distances between any two grid nodes. Let (g_{x_i}, g_{y_i}) be the grid coordinate of node $s \in S'$. Then, the distance between two grid nodes s_i and s_j in the established grid coordinate system is

$$D_{ij} = \sqrt{(g_{x_i} - g_{x_j})^2 + (g_{y_i} - g_{y_j})^2} \quad (1)$$

Without loss of generality, assuming that $|S'| = M$ and $S' = \{s_i \| i = 1, 2, ..., M\}$. The error function is defined as.

$$\epsilon(X) = \sum_{i=1}^{M} \sum_{j=1}^{M} (D_{i,j} - \tilde{d}_{i,j})^2 \quad (2)$$

Where $X = \{(g_{x_i}, g_{y_i}) \| i = 1, 2, ..., M\}$. A PSO algorithm is applied to minimize the error function and give each node the grid coordinate. PSO is an optimization technique developed by Kennedy and Eberhart in 1995, inspired by social behavior of organisms such as bird flocking and fish schooling [11]–[13]. PSO is a population-based search method where individuals, called particles, change their position with time. Particles change their position by flying around in a multidimensional problem space. During flight, each particle keeps track of its coordinate that has the best solution encountered by itself so far and modifies its position based on its own history and the history of all other particles to make use of the best position. Therefore, PSO combines local search methods with global search methods. Here, each particle represents a solution set of grid coordinates of nodes in S'. Each particle evaluates how good the solution is according to the value, called fitness, of the error objective function in (2). One of the advantages of the proposed grid-based localization approach is that the PSO algorithm can be calculated efficiently because the number of grid nodes used to create the backbone of grid coordinate system is usually small (e.g., 5 to 15 grid nodes). Specifically, the number of grid nodes can be adjusted adaptively by setting the grid width according to the requirements of target applications. Furthermore, the grid-based coordinate system digitizes the solution space of the PSO algorithm. Let $K = (L/w)^2$ be the number of grid points in the established grid coordinate system. The upper bound of position vectors needed to be checked is

$$C_M^K \cdot M! = \prod_{i=1}^{M} (K - i + 1) \quad (3)$$

 ISBN: 978-1-61208-296-7

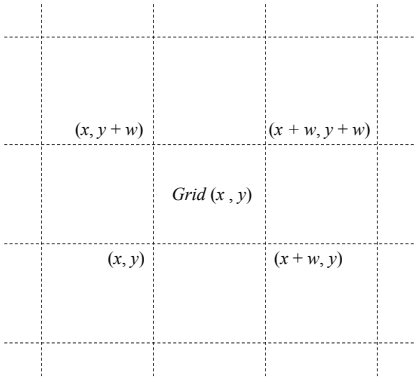


Fig. 2. Grid(x, y).

Fig. 1(a)-(d) illustrate an example of the grid coordinate system establishment in the network consisting of ten nodes. Fig. 1(a) is the constructed graph $G(V, E)$, where $V = \{01, 02, ..., 10\}$ and $E = \{(01, 04), (01, 05), (02, 04), (03, 05), (03, 06), (04, 05), (05, 06), (05, 08), (06, 07), (06, 10), (07, 08), (07, 09), (07, 10)\}$.
Fig. 1(b) shows the set of grid nodes $S' = \{02, 05, 09, 10\}$ for establishing the grid coordinate system.
$\epsilon = (d_{02,05} - \tilde{d}_{02,05})^2 + (d_{02,09} - \tilde{d}_{02,09})^2 + (d_{02,10} - \tilde{d}_{02,10})^2 + (d_{05,09} - \tilde{d}_{05,09})^2 + (d_{05,10} - \tilde{d}_{05,10})^2 + (d_{09,10} - \tilde{d}_{09,10})^2$
Fig. 1(c) shows the established local grid coordinate system. Note that the coordinate system is relative coordinate system and it can be arbitrary rotated or translated as long as the relative distances between nodes remain unchanged. To facilitate the comparison between the real positions of grid nodes and the established grid coordinates, the coordinate system in Fig. 1(c) is transformed to obtain Fig. 1(d). The node with lowest ID is set to be the origin $(0, 0)$, the node with the second lowest ID is set to be on positive X-axis with coordinate $(x, 0)$ where $x > 0$, and the node with the third lowest ID is set to be on positive Y-axis. In 1(d), the small circles ("o") denote the real coordinates of the nodes and the squares ("$*$") represent the grid coordinates after coordinate transformation.

B. The Non-grid Nodes

Let $Grid(x, y)$ denote the grid shown in Fig. 2. Then, $(x + w/2, y + w/2)$ is the center coordinate of $Grid(x, y)$. $Grid(x, y)$ is said to be within the $R + \epsilon$ range of the grid node s_i if $\sqrt{(x + w/2 - g_{x_i})^2 + (w + w/2 - g_{y_i})^2} \leq R + \epsilon$. After the grid coordinate system is established, the non-grid nodes compute their locations, i.e., the grid they are located in, by sending a request message to the nearby grid nodes (within the transmission range R) on demand. Let G_i be the set of grid nodes that receive the request message from non-grid node s_i. Upon receiving the request, every grid node $s_j \in G_i$ responses a message containing its identity number, the grid coordinate (g_{x_i}, g_{y_i}), and the set H_j of grid coordinates that are within the $R + \epsilon$ range of grid node s_j. (ϵ is a tunable parameter used to tolerate possible measurement errors in estimation of distance.) Let $H = \cap_{s_j \in G_i} H_j$ be the intersection set of grid coordinates. With the received information from nearby grid nodes, s_i calculates its own location (p, q) by calculating the following formula.

$$(p, q) = \arg \min_{(x,y) \in H} \{f(x, y)\}$$

where

$$f(x, y) = \sum_{s_k \in G_i} (\sqrt{(x + w/2 - g_{x_k})^2 + (y + w/2 - g_{y_k})^2} - \tilde{d}_{i,k})^2 \quad (4)$$

The proposed localization scheme for the non-grid nodes is a digitized triangulation approach. The advantage of this scheme is that every non-grid node can be positioned efficiently because the total number of grids needed to be evaluated in (4) is bounded by $(\frac{(R+\epsilon)^2 \cdot \pi}{w^2})$. Since R and w are usually constant in most wireless sensor networks, the computation complexity of the localization approach for non-grid nodes is $O(1)$. For example, let $R = 40, w = 15$ and $\epsilon = 5$ in a wireless sensor networks. Then, each regular node only requires evaluating at most 28 grids to determine the best grid in which they reside. Fig. 3 shows an example to illustrate this scheme more clearly. Assume $G_i = \{A, B, C\}, R + \epsilon = 3, w = 1, \tilde{d}_{i,A} = 1, \tilde{d}_{i,B} = 2$, and $\tilde{d}_{i,C} = 3$. $H = \{(1, 1), (2, 1), (3, 1), (1, 2), (2, 2), (3, 2), (2, 3)$. $f(2, 2) \simeq 3.3$.

IV. Simulation

In this section, we evaluate the performance of the proposed localization technique via extensive simulations. Particularly, we study the impact of several parameters of interest, such as the number of grid nodes, grid width, and node density (grid nodes plus non-grid nodes). The proposed grid-based localization method was implemented in Matlab. Measurement error in the estimation of distances between adjacent nodes were modeled as zero-mean additive white Gaussian noise, which was also adopted in [17]. In the following evaluations, the sensor nodes are randomly deployed in a 100×100 square area, i.e., $L = 100$.

A. Grid Nodes

1) Grid Width versus Number of Grid Nodes: Extensive simulations have shown that the more distance measurements used to locate nodes, the higher accuracy of the established grid coordinate system, since the errors in the estimation of distances could be smoothed out more effectively. However, since the coordinate of every grid node is restricted by grid coordinate (see the error function in (2)), this would bring additional errors slightly, depending on the grid width and the total number of grid nodes used in the establishment of grid coordinate system. To evaluate the accuracy of the established grid coordinate system, the grid coordinate system is transformed by setting the node with lowest ID to be the origin $(0, 0)$, the node with the second lowest ID to be on positive X-axis with coordinate $(x, 0)$ where $x > 0$, and the node with the third lowest ID to be on positive Y-axis. The average error of the grid coordinate system is evaluated as follows.

$$\psi = \frac{\sqrt{\sum_{i=1}^{M} (x_i - g_{x_i})^2 + (y_i - g_{y_i})^2}}{M} \quad (5)$$

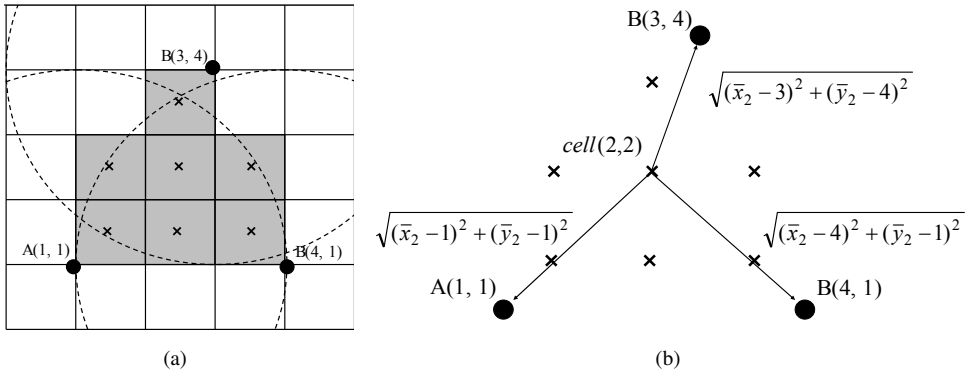


Fig. 3. Location computation of non-grid nodes. (a) The intersection set. (b) The distances between grid cell (2,2) and grid nodes.

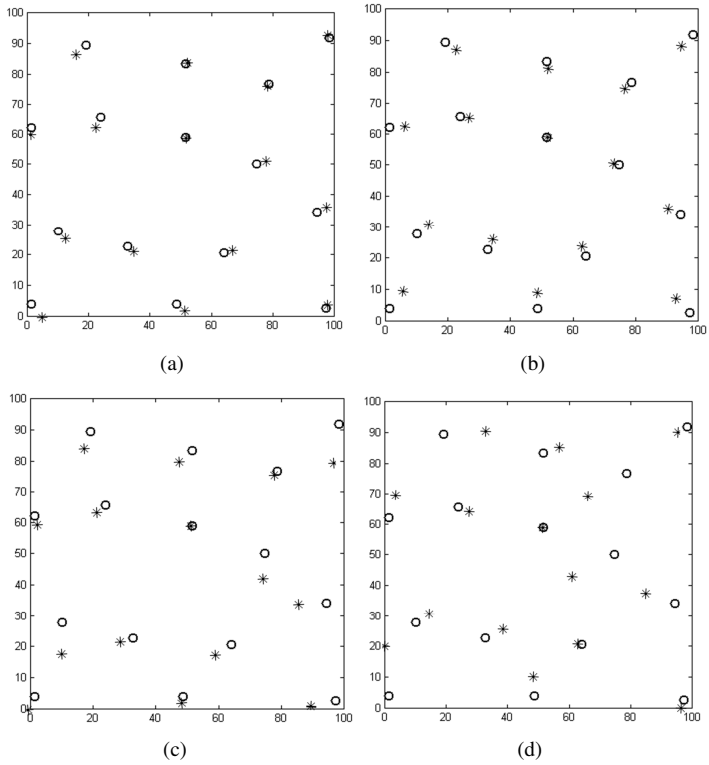


Fig. 5. The established grid coordinate system with $M = 15$. (a) $\psi = 1.7804, w = 5$. (b) $\psi = 4.0752, w = 10$. (c) $\psi = 4.8085, w = 15$. (d) $\psi = 8.4885, w = 20$.

where (x_i, y_i) is the real coordinate of grid node s_i and (g_{x_i}, g_{y_i}) is the calculated grid coordinate after coordinate transformation, and M is the total number of grid nodes. Here, ψ represents the average of distance between the real coordinates of grid nodes and the grid coordinates. Fig. 4 shows the average error of the grid coordinate system where the X-axis represents grid width w and the Y-axis denotes the average error ψ. As demonstrated in Fig. 4, ψ increases (approximately) linearly with grid width w and the average error is no more than $w/2$ when $M \leq 20$ and $1 \leq w \leq 20$. (Note that $M \leq (100/w)^2$ in a 100×100 area.) Additionally, ψ increases about $k \cdot (M_2 - M_1) \cdot w$ (e.g., in our simulations, $k \simeq 1/50$) as M increases from M_1 to M_2. For example, ψ increases about 3 units when M changes from 10 to 20 and $w = 15$.

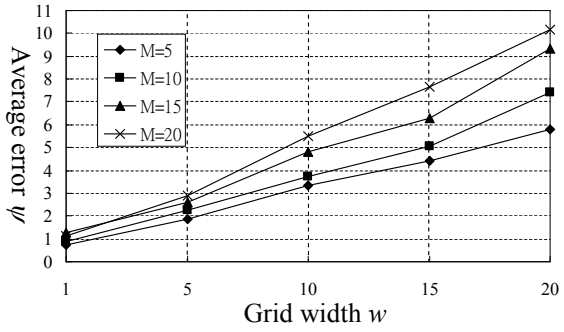


Fig. 4. Average grid coordinate error versus grid width.

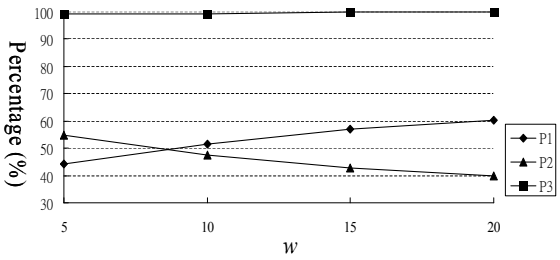


Fig. 6. The localization accuracy of non-grid nodes versus grid width.

2) Grid Coordinate System: Fig. 5(a)-(d) depicts the established grid coordinate system with $M = 15$. The small circles ("o") denote the real coordinates of the nodes and the squares ("$*$") represent the calculated grid coordinates after coordinate transformation. As shown in Fig. 5, the real coordinates are very close to the estimated coordinates when $w = 5$ and 10. As w increases, the differences between them become slightly obvious, but the simulation results in Fig. 4 demonstrate that the average error is no more than 10 units. This simulation also verifies that our proposed grid-based localization indeed establishes the local grid coordinate system.

B. Non-grid Nodes

We simulate networks containing 100 randomly distributed sensor nodes. Fifteen nodes are selected to serve as grid nodes, i.e., $M = 15$, and the non-grid nodes communicate with the nearby grid nodes to obtain necessary information. Then, they determine which grid they belong to using (4). We set $R = 40$ units so that each non-grid node has an average of about five neighboring grid nodes. The non-grid nodes that their real positions are near grid boundary in the grid coordinate system may be located incorrectly due to errors of the estimated distances between non-grid nodes and grid nodes. Let (x_i, y_i) be the real grid coordinate in which node s_i is located and $(\tilde{x}_i, \tilde{y}_i)$ be the estimated grid coordinate, respectively. Let P_1 represent the percentage of non-grid nodes that locate themselves correctly, i.e., $(\tilde{x}_i, \tilde{y}_i) = (x_i, y_i)$. Let P_2 denote the percentage of non-grid nodes that locate themselves *partially correct*. The localization result of a node s_i is said to be partially correct if $(\tilde{x}_i, \tilde{y}_i)$ equals to one of eight grids surrounding the real grid (x_i, y_i). Let $P_3 = P_1 + P_2$. Fig. 6 demonstrates that P_1 increases slightly with grid width w. About sixty percent of non-grid nodes locate themselves at correct grid as $w \leq 20$. Significantly, although some of non-grid nodes (near 40%) locate themselves incorrectly, almost all of them are located correctly in one of eight grids surrounding their real positions.

V. CONCLUSION

In this paper, we propose a scalable grid-based localization technique to balance the tradeoff between computation complexity and precision of localization. The simulation results show that a small number of nodes, e.g., five to fifteen, are usually enough for create the coordinate system with acceptable localization accuracy for many applications. The average error of the grid coordinate system is no more than half the grid width. About sixty percent of non-grid nodes could compute in which grids they are located correctly. Significantly, although some of non-grid nodes (near 40%) locate themselves incorrectly, almost all of them are located correctly in one of eight grids surrounding their real positions.

ACKNOWLEDGMENT

This research was partially supported by NSC under Grant No. 100-2218-E-006-035-MY2 and No. 101-2221-E-006-235-.

REFERENCES

[1] I. F. Akyildiz,W. Su, Y. Sankarasubramaniam, and E. Cayirci, Wireless sensor networks: A survey, Computer Networks, vol. 38, no. 4, pp. 393-422, March 2002.

[2] D. Estrin, D. Culler, K. Pister, and G. Sukhatme, "Connecting the Physical World with Pervasive Networks," IEEE Pervasive Computing, vol. 1, no. 1, pp. 59-69, January-March 2002.

[3] X. Li, J. Yang, A. Nayak, and I. Stojmenovic, "Localized Geographic Routing to a Mobile Sink with Guaranteed Delivery in Sensor Networks," IEEE Journal on Selected Areas in Communications, vol. 30, no. 9, pp. 1719-1729, Oct. 2012.

[4] A. Awad, R. German, and F. Dressler, "Exploiting Virtual Coordinates for Improved Routing Performance in Sensor Networks," IEEE Trans. on Mobile Computing, vol. 10, no. 9, pp. 1214-1226, Sep. 2011.

[5] B. Li, W. Wang, Q. Yin, H. Li, and H.-M. Wang, "Energy-efficient cooperative geographic routing in wireless sensor networks," IEEE Int'l Conf. Communications (ICC), June 2012, pp. 152-156.

[6] G. Tan and A.-M. Kermarrec, "Greedy Geographic Routing in Large-Scale Sensor Networks: A Minimum Network Decomposition Approach," IEEE/ACM Trans. on Networking, vol. 20, no. 3, pp. 864-877, June 2012.

[7] D. Nicolescu and B. Nath, DV-based positioning in ad hoc networks, Journal of Telecommunication Systems, vol. 22, no. 1-4, pp. 267-280, 2003.

[8] S. Zhu and Z. Ding, "Distributed Cooperative Localization of Wireless Sensor Networks with Convex Hull Constraint," IEEE Trans. on Wireless Communications, , vol. 10, no. 7, pp. 2150-2161, July 2011.

[9] J. Gu, S. Chen, and T. Sun, "Localization with Incompletely Paired Data in Complex Wireless Sensor Network," IEEE Trans. on Wireless Communications, vol. 10, no. 9, pp. 2841-2849, Sep. 2011.

[10] R. Sugihara and R. K. Gupta, "Sensor localization with deterministic accuracy guarantee," IEEE INFOCOM, April 2011, pp. 1772-1780.

[11] J. Kennedy and R. Eberhart, "Solution Swarm Optimization," IEEE Int'l Conf. Neural Networks, 1995, pp. 1942-1948.

[12] R. Eberhart and J. Kennedy, A new optimizer using solution swarm theory, Int'l Symp. Micro Machine and Human Science, 1995, pp. 39-43.

[13] R. Eberhart and Y. Shi, Comparison between Genetic Algorithms and Particle Swarm Optimization, Int'l Conf. Evolutionary Programming, 1998.

[14] M. Leng, W.-P. Tay, and T.Q.S. Quek, "Cooperative and distributed localization for wireless sensor networks in multipath environments," IEEE Int'l Conf. Acoustics, Speech and Signal Processing (ICASSP), March 2012, pp. 3125-3128.

[15] B. Zhang, X. Cheng, N. Zhang, Y. Cui, Y. Li, and Q. Liang, "Sparse target counting and localization in sensor networks based on compressive sensing," IEEE INFOCOM, April 2011, pp. 2255-2263.

[16] K.-S. Low, H.A. Nguyen, and H. Guo, "A particle swarm optimization approach for the localization of a wireless sensor network," IEEE Int'l Symp. Industrial Electronics (ISIE), 2008, pp. 1820-1825.

[17] H. Wu, C. Wang, and N. F. Tzeng, "Novel self-configurable positioning technique for multi-hop wireless networks, IEEE/ACM Trans. on Netowkring, vol. 13, no. 3, pp. 609-621, June 2005.

[18] M. Luby, A simple parallel algorithm for the maximal independent set problem, SIAM Journal on Computing, vol. 15, no. 4, pp. 1036-1055, Nov. 1986.

 ISBN: 978-1-61208-296-7

Implementation of Controlled Sink Mobility Strategies with a Gradient Field in Wireless Sensor Networks

Shinya Toyonaga*, Yuki Fujita*, Daichi Kominami† and Masayuki Murata*
*Graduate School of Information Science and Technology, Osaka University, Osaka, Japan
Email:{s-toyonaga, y-fujita, murata}@ist.osaka-u.ac.jp
†Graduate School of Economics, Osaka University, Osaka, Japan
Email:d-kominami@econ.osaka-u.ac.jp

***Abstract*—Reduced energy consumption and extension of network lifetime are important challenges for wireless sensor networks, due to the energy-constraint properties of its elements. To solve these problems, many studies have been conducted on incorporating mobile sinks and controlled mobility into wireless sensor networks. When a mobile sink relays data from a sensor network to an operator, some sensors can save energy by reducing the number of transmitted packets and the communication range. Another important advantage is that sparse and disconnected networks are better handled when a mobile node relays packets between networks. In existing methods, however, all or some nodes must know their own location through the use of devices such as Global Positioning System (GPS) receivers. Yet it entails a high cost when each node is equipped with a GPS receiver. Moreover, GPS-based localization solutions cannot provide reliable location estimate in indoor environments, and in the presence of obstacles. We propose a method for controlling the mobility of a mobile sink so that it moves toward a targeted sensor node without GPS. In the proposed method, a sensor node selected as a target node broadcasts a control message, and each node records a hop count from the target node so as to construct a gradient field. A mobile sink can approach the target node using the gradient field. We evaluate our method by computer simulation and experiments using a cleaning robot that implements our method. We show that a mobile node with our proposed method can reach a target node in about 6 min.**

***Index Terms*—mobile sink; controlled mobility; sensor network.**

I. Introduction

The main challenges for Wireless Sensor Networks (WSNs) include reduced energy consumption and lifetime extension. To solve these problems, many controlled mobility protocols have been studied [1-6]. A mobile node that moves to collect information in WSNs is called a mobile sink. In controlled mobility, an operator deploys mobile sinks, whose mobility is dynamically controlled by received information from inside or outside of networks.

When a mobile sink relays data from sensor networks to an operator, some sensors can save energy from the reduced number of transmitted packets and communication range. Another important advantage is that sparse, disconnected networks can be better handled, since a mobile node can relay packets from one network to another. However, in existing methods, it is important to point out that all or some nodes need to know their own location information using devices such as GPS receivers. Yet, it entails a high cost when each node is equipped with a GPS receiver. Moreover, GPS-based localization solutions cannot provide reliable location estimate in indoor environments, and in the presence of obstacles.

In this paper, we propose a method to control the mobility of a mobile sink so that it moves toward a target sensor node by using a gradient field. The concept of a gradient field has been applied in gradient-based routing [7]. In gradient-based routing, all nodes have a gradient, which is a value presenting the direction through which the sink can be reached. Such a gradient field can be set up according to different information, such as hop count, energy consumption, or physical distance. When each node forwards a data along the gradient field, the data can reach the sink. In this paper, we use a gradient field for guiding a mobile sink to a location where a specific node is deployed. For example, a device such as a Personal Digital Assistant (PDA) floods a message, and PDAs that receive the message record the hop count from the flooding PDA. When such a hop count is assigned to PDAs, they can guide a mobile sink, such as a rescue robot, to the position where a person is requesting help.

In the proposed method, a sensor node selected as a target node floods a control message, and each node records a hop count from the target node. Then, a gradient field towards the target node is set up. When a mobile sink moves along the gradient field, like a data flow in gradient-based routing, it can reach the target node. For that, a mobile sink intercepts hop count information exchanged in the network. It then approaches the sensor node with smaller hop count and finally reaches the target node.

In our method, a mobile sink does not use any GPS receivers, but rather a Received Signal Strength Indication (RSSI) of messages exchanged in the sensor network. Because a gradient field is insufficient for a mobile sink to determine the direction to the node, mobile sinks measure the RSSI of a message from a sensor node, and searches for directions in which the RSSI is larger. Specifically, if a mobile sink goes straight and the RSSI increases, it continues to go straight because it approaches the sensor node. When going straight decreases the RSSI, the mobile sink measures RSSI while turning, and heads in the direction in which the RSSI is largest. By repeating these processes, a mobile sink can get closer to

 ISBN: 978-1-61208-296-7

TABLE I
DESCRIPTION OF CONTROL MESSAGES

Information	Description
node_id	Sender's node ID
hop_count	Hop count from sender to target
net_id	Sender's network ID

the sensor node.

The main advantage of using a gradient field is that a mobile sink needs not know the global information of the whole network. Instead, it only needs to know the local gradient information. Therefore, all nodes do not have to communicate over a long distance nor to send much information about a network.

We evaluate our proposed method by computer simulation and experiment using a cleaning robot. Our simulation and experiment results show that a mobile sink can be guided to the target node by using a gradient field. Since the main scope of our method in this paper is the mobility control of the mobile sink, our simulation and experiment focus on the capability of the mobile sink to be guided toward the target node by using a gradient field.

The rest of this paper is organized as follows: Section II presents our proposed method of controlled mobility. In Section III, we validate our method by computer simulation. Section IV discusses implementation using a cleaning robot, and we evaluate our method using the cleaning robot and sensors in Section V. Finally, Section VI presents our conclusions.

II. CONTROLLED MOBILITY WITH A GRADIENT FIELD

In our protocol, each node has a hop count from a target node, which is regarded as a gradient field. To assign a hop count from the target node to each node and update it, all sensor nodes send a control message to their neighbor nodes. A target node assigns its hop count to zero and generates a control message containing hop count that is initially set to one and advanced in increments by the forwarding nodes. The other node broadcasts the control message and assigns its hop count to the hop count contained in it. If a node receives more than one control message, the node assigns the hop count to the smallest hop count and broadcasts the message containing the smallest hop count. Table I shows the information included in the control message. Each node is assumed to have a unique identifier (node ID) to break ties, which can be easily removed by allowing each node to choose a random number in a large enough interval. Network ID is an identifier of the network which is comprised of a node connected to the target node. Mobile sinks intercept this control message to control their own mobility.

When a mobile sink receives a control message, it starts to control its own mobility along the gradient of the hop count field. First, it checks whether the network ID of the control message is the same as one it recorded. When the network ID is different from the recorded one, it ignores the control message. Otherwise, a mobile sink checks the hop count of the control message, and compares it to the hop count which it has recorded as the smallest one. If the received hop count is smaller than the currently recorded one, the mobile sink records the sender's node ID and its hop count, then tries to approach the sender. After the mobile sink receives data from a target node, it needs to restore its mobile algorithm because it may deliver data to a base station or move to the next network in which the other target node is located. When a mobile sink receives the data from a target node, it therefore records the network ID as the node that it has been arrived at and ignores control messages containing the same network ID as the recorded one.

However, a mobile sink cannot determine the direction of the sensor node because of feature of non-directional radio. We thus propose a controlled mobility using an RSSI. Our proposed method is applicable when mobile sinks are equipped with a non-directional antenna, but it can determine the direction more precisely when equipped with a directional antenna. Figure 1 shows the algorithm, which describes the operation of the controlled mobility. In that algorithm, h_{min} is a hop count, node_{min} is a node ID, and RSSI_{min} is an RSSI, when a mobile sink receives a control message containing a hop count that is the smallest it has received. net_{id} is the network ID of the target node that a mobile sink needs to approach to currently.

When a mobile sink receives a control message, it checks whether the sender is a target node (lines 1 and 2). If so, the mobile sink communicates with the target node and restores its own mobile algorithm (lines 3–6). Otherwise, the mobile sink checks the network ID included in the control message whether it is the same as net_{id}. If the network ID is different from net_{id}, a mobile sink ignores the control message (lines 8 and 9). Otherwise, the mobile sink checks the hop count included in the control message. If the received hop count is smaller than h_{min}, the mobile sink updates h_{min}, node_{min}, and RSSI_{min}, because it needs to approach the sender (lines 11–14). The mobile sink then searches for the direction to the sender (line 15). The operation of the searching direction to the node is described in Figure 2. If the received hop count is larger than h_{min}, the mobile sink ignores the control message (lines 16–18). If the received hop count is the same as h_{min}, the mobile sink checks the node ID included in the control message. If the node ID of the sender is different from node_{min}, the mobile sink ignores the control message and goes straight (lines 19 and 20). If the node ID of the sender is the same as node_{min}, the mobile sink checks the RSSI included in the control message. If the received RSSI is larger than RSSI_{min}, the mobile sink continues to go straight because it is approaching the sender (lines 21–23). Otherwise, the mobile sink searches for the direction to the sender, because it is getting farther from the sender (line 25).

In the following section, we use computer simulation to show that the proposed method can realize controlled mobility.

III. VERIFICATION OF CONTROLLED MOBILITY BY COMPUTER SIMULATION

We evaluate our method by computer simulation to show that the mobility of a mobile sink can be controlled. The monitoring area is 120×80 m. Seven sensor nodes and one

Algorithm 1 A mobile sink approaches a target node

```
receive message m
if the mobile sink receives m from target node then
    the mobile sink receives data from target node
    the mobile sink records m.net_id
    the mobile sink moves according to the usual mobile algorithm
    return
end if
if net_id ≠ m.net_id then
    return
end if
if h_min > m.hop then
    h_min ⇐ m.hop
    node_min ⇐ m.node_id
    RSSI_min ⇐ m.RSSI
    search direction
else if h_min < m.hop then
    Go straight
else
    if node_min ≠ m.node_id then
        Go straight
    else if RSSI_min < m.RSSI then
        RSSI_min ⇐ m.RSSI
        Go straight
    else
        search direction
    end if
end if
```

Fig. 1. Approach algorithm toward a target node

Algorithm 2 A mobile sink searches for the direction to the sender

1: The mobile sink moves on the circumference centering on the current position (L_c) and records the position (L_max) where the mobile sink receives the message with the maximum RSSI.
2: The mobile sink returns to (L_c).
3: The mobile sink goes straight in the direction toward position (L_max).

Fig. 2. Direction search algorithm toward the sender

mobile sink are deployed at random places, and one of seven sensor nodes behaves as a target node. The model of radio attenuation is the free space model [8] and we assumed that no noise exists.

Our method is implemented in the OMNeT++ 4.1 [9] network simulator. We evaluate T_{move}, which is the time required for the mobile sink to approach the target node and to receive a control message from the target node. The default mobility algorithm of the mobile sink is the random way point model [10], which switches to the mobility algorithm shown in Section II, when it receives a control message from a sensor node. A mobile sink is deployed at random place in initial settings. The mobility based on the random way point model is shown in following.

1) A mobile sink selects the random place in the monitoring area.
2) It goes straight towards the selected place at constant velocity.

TABLE II
SIMULATION CONFIGURATION

Parameter	Value
Control message send time	1 s
Communication range	50 m
Data packet size	128 byte
Bandwidth	250 kbps
Mobile sink velocity	1 m/s
Mobile sink angular velocity	$\frac{\pi}{6}$ rad/s

TABLE III
SIMULATION RESULTS

	T_{move}
with controlled mobility	138.78 ± 67.17 s
without controlled mobility	308.30 ± 102.58 s

3) When it reaches the selected place, its mobility process returns to 1).

When a mobile sink goes straight and hits against a boundary of a monitoring area, it goes straight to the reflection direction. When a mobile sink searches direction to the node, it moves in a circle. When it moves in a circle and hits against a boundary of a monitoring area, it turns around in reverse. The details of the simulation configuration are summarized in Table II.

To compare the influence of our method, we evaluate T_{move} when the mobile sink moves only according to the random way point model. Table III shows simulation results. The number of trials is 50, and the confidence interval is 95%.

T_{move} is smaller when our method is implemented, so controlled mobility using a gradient field can be realized. The confidence interval when not using our method is larger due to the randomness of the random way point model, but the ratio between the confidence interval and the average T_{move} is larger under our method. This is because the number of circular movements when searching for a direction varies considerably according to the location where the mobile sink first receives a control message.

IV. IMPLEMENTATION OF CONTROLLED MOBILITY USING A GRADIENT FIELD

In this section, we describe how to implement controlled mobility in a mobile sink. We present an outline, and then details regarding configuration and system implementation.

A. Outline

For method implementation, as a mobile sink we use a patrolling robot such as an automatic cleaning robot. When the cleaning robot receives a control message from a sensor node while cleaning, it is guided to a target node along the gradient field. In our experimentation, the robot restores its mobility for cleaning after receiving a control message from the target node.

B. Configuration

In our experiment, we use a Roomba 790 (iRobot Corp.), an automatic cleaning robot with a publicly documented serial interface [11]. For sensor nodes we use the IRIS Mote XM2100 (Crossbow Technology) [12]. IRIS wireless modules

 ISBN: 978-1-61208-296-7

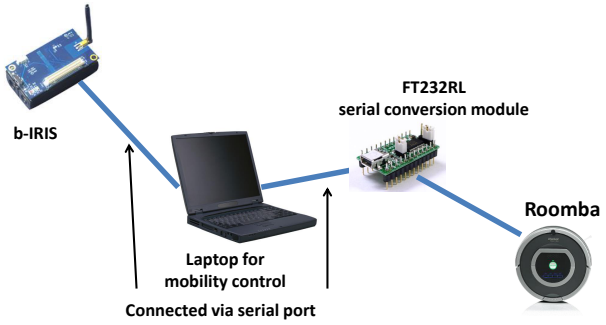


Fig. 3. Components of the mobile sink

TABLE IV
ROOMBA MOBILE PHASES

Phase	Description
FIND_NEWIRIS	Find an s-IRIS with smaller hop count
APPROACH_IRIS	Approach the s-IRIS with recorded node ID
REACH_TARGET	Reach target node
CLEAN	Do not control mobility and move based on default mobile strategy

are widely used for WSN development, are IEEE 802.15.4 compliant, and can be programmed using C#, Java, and other languages. We implement two types of IRIS with different functions. One is an s-IRIS (sensing IRIS), which generates sensing data and constructs the gradient field. The other is a b-IRIS (base station IRIS), which intercepts control messages and RSSI, and forwards them to the Roomba's mobility controller. We prepare a laptop computer for controlling the mobility of Roomba and attach it on the Roomba. The laptop computer decides the strategy of the mobility based on received information.

We used devices connected as shown in Figure 3. The b-IRIS was connected with a laptop computer via a serial port, and the laptop was connected with the Roomba via a serial conversion module (FT232RL). The laptop sends commands to the Roomba via the FT232RL, which serializes the command and forwards the results.

C. Implementation

Our proposed controlled mobility consists of three phases. The b-IRIS selects the next mobile phase based on information received from an s-IRIS. The b-IRIS then sends the next mobile phase and RSSI of the received control message to the Roomba's mobility controller, which selects an appropriate command according to the mobile phase and RSSI, and sends the command to the Roomba. Table IV shows these phases. In the following, we describe the processes and implementation of the b-IRIS and mobility controller.

1) B-IRIS processes : Figure 4 shows the state transition diagrams of the b-IRIS's process flow. The b-IRIS manages the mobile phase, as shown below, and informs the mobility controller of the next mobile phase. h_{min} is a hop count, node_{min} is a node ID, and RSSI_{min} is an RSSI, when the b-IRIS receives a control message containing the hop count that is the smallest so far received.

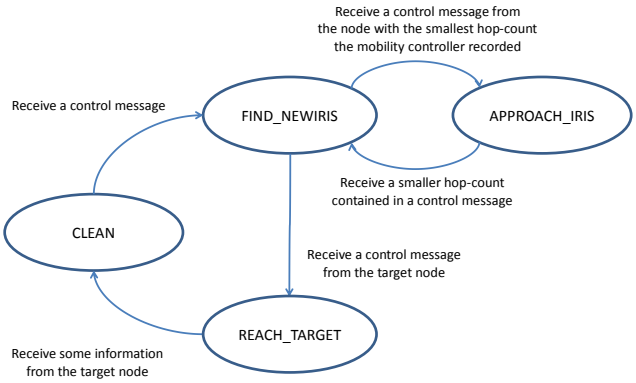


Fig. 4. State transition diagrams of the Roomba's mobile phases

a) CLEAN: When the b-IRIS receives a control message from an s-IRIS, it enters the FIND_NEWIRIS mobile phase.

b) FIND_NEWIRIS: When the b-IRIS receives a control message from a target node, it records the network ID to which the target node belongs, then enters the REACH_TARGET mobile phase. When the b-IRIS receives a control message from an s-IRIS, it informs the mobility controller of RSSI, and then the next mobile phase is APPROACH_IRIS.

c) APPROACH_IRIS: When the b-IRIS receives a control message from an s-IRIS, it compares the received hop count with h_{min}.

- If the received hop count is smaller than h_{min}, the b-IRIS updates h_{min} and node_{min}, then enters the FIND_NEWIRIS mobile phase.
- If the received hop count is larger than h_{min}, the b-IRIS does nothing.
- Otherwise, the b-IRIS checks the node ID of the sender. If the sender's node ID is the same as node_{min}, the b-IRIS informs the RSSI mobility controller. Otherwise, the b-IRIS records the node ID of the larger of the received RSSI and RSSI_{min}.

d) REACH_TARGET: When the b-IRIS receives a control message from a target node, it checks the received network ID. If the ID is already recorded, it ignores the received control message for a while. The next mobile phase is CLEAN.

2) Mobility controller processes: The mobility controller commands the Roomba based on the mobile phase and RSSI received from the b-IRIS.

a) CLEAN: When the mobility controller commands the Roomba to clean, the Roomba moves according to its original mobility algorithm.

b) FIND_NEWIRIS: The mobility controller discards and initializes all saved data.

c) APPROACH_IRIS: The Roomba approaches an s-IRIS. When the mobility controller searches for the direction of the s-IRIS, it commands the Roomba to rotate. The mobility controller searches for the direction with the largest RSSI, and goes straight in that direction. In order to exaggerate the difference in RSSI values, the b-IRIS is equipped with a pseudo-directional antenna in our experiments. The outline of the directional searching procedure is described below.

 ISBN: 978-1-61208-296-7

1) While the Roomba rotates, the mobility controller records the direction with the largest RSSI. When the mobility controller receives equally large RSSI from multiple directions, the mobility controller records the composition of those directions.
2) After the Roomba turns 360 degrees, the mobility controller commands the Roomba to turn to the recorded direction.
3) The mobility controller commands the Roomba to go straight. If the received RSSI tends to decrease while Roomba goes straight, the mobility controller processes 1).

In the following, we describe the directional searching procedure in detail. Each s-IRIS sends its hop count every second, and the angular velocity of Roomba is $\frac{2\pi}{N}$ rad/s when Roomba receives N messages and RSSI while it turns 360 degrees. The mobility controller numbers directions from 0 to $N-1$. This number is denoted by $RoombaDir$, and is initialized as 0 when the Roomba starts to turn. Then, every time the mobility controller receives RSSI, $RoombaDir$ is incremented through each direction. The mobility controller records the $RoombaDir$ with the largest RSSI as $TopDir$. If there are multiple directions with equally large RSSI, the mobility controller records the last two $RoombaDir$ values as $RD1$ and $RD2$, and the mobility controller records $\frac{RD1+RD2}{2}$ as $TopDir$. When the difference between $RD1$ and $RD2$ is $\frac{N}{2}$ or more, the mobility controller records $\frac{RD1+N+RD2}{2} \bmod N$ as $TopDir$ so that the Roomba does not go just in the opposite direction. The mobility controller uses an N s timer to decide when to finish searching directions. In an actual environment, the b-IRIS cannot always receive hop counts. Therefore, the mobility controller sets RSSI of the direction to 0 using the timer when the b-IRIS does not receive a hop count.

When the mobility controller decides $TopDir$, it needs to command Roomba to face the direction $TopDir$. When $TopDir$ is $\frac{N}{2}$ or more, the mobility controller commands the Roomba to turn for $(N - TopDir)$ seconds clockwise. Otherwise, the mobility controller commands the Roomba to turn for $(TopDir)$ seconds counterclockwise.

In an actual environment, the mobility controller may happen to go to the wrong direction because of noise or radio interference. Therefore, when Roomba goes on wrong direction, the mobility controller needs to revise the direction of movement of Roomba. In order to do so, the mobility controller determines the increase and decrease of RSSI tendency by the exponential moving average (EMA) of the RSSI. The EMA of the RSSI is calculated by equation (1). Every T seconds, the mobility controller calculates the EMA of RSSI and compares it with that of T seconds ago. E_n and R_n describe the EMA and RSSI respectively when the mobility controller observes the nth RSSI. α is a smoothing coefficient.

$$E_n = (1-\alpha) \cdot E_{n-1} + \alpha \cdot R_n \quad (1)$$

In this experiment, E_1 is equal to R_1. The mobility controller compares E_n with E_{n-5}. It determines that RSSI is increasing when E_n is larger than E_{n-5}, and decreasing otherwise.

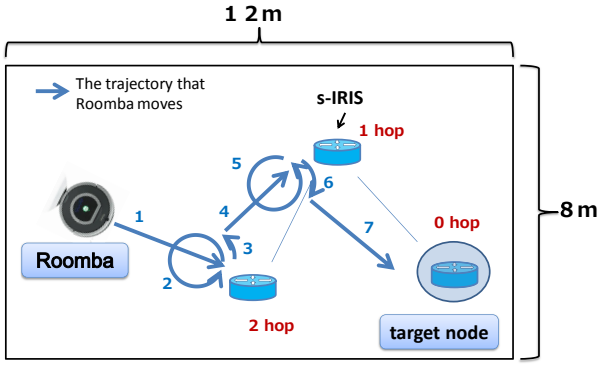


Fig. 5. The deployment of s-IRISs and the mobile sink

d) REACH_TARGET: The mobility controller commands the Roomba to stop for enough time to receive a control message from the target node.

V. Experimental Evaluation

In this section, we show that the Roomba's mobility can be controlled by our proposed method. To show that, we compare the time taken for the Roomba to approach a target node T_{move} when our method is implemented and when it is not. To approach the target node means that the b-IRIS receives a message with RSSI larger than a predefined threshold from the target node. IRIS operates under IEEE 802.15.4 in non-beacon mode, and can asynchronously communicate with other IRIS modules at all times. In this experiment, N is set to 14, α is set to 0.2 and T is set to 5.

A. Outline of experiments

Each IRIS is deployed in a room to construct a multi-hop network. To show the influence of controlled mobility, a single IRIS behaves as the target node. We use the topology shown in Figure 5. The target node IRIS is located at the network boundary. The Roomba starts to move from a specific location, and we measure T_{move}. To measure the efficiency of our algorithm, we compare T_{move} when the Roomba moves along the gradient field and T_{move} when the Roomba moves according to its default algorithm.

B. Experimental environment

The dimensions of the room used for the experiment was 12×8 m. Three IRIS modules were deployed in the room. We removed obstacles from the room to avoid impeding the Roomba's movement. The WSN bandwidth was 2.4 GHz, which overlapped with the bandwidth used for a wireless LAN. We used the 26th channel (2480 MHz), which investigation showed to have the least noise.

In the experiment, the Roomba starts to move after the gradient field is constructed in the WSN. The initial position of the Roomba is the same in each experiment, and experiments start by running a program implemented on the laptop. Figure 6 shows the process flow.

 ISBN: 978-1-61208-296-7

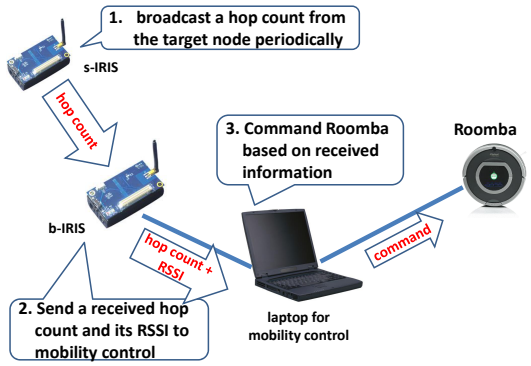


Fig. 6. The process flow

TABLE V
EXPERIMENTAL RESULTS

	T_{move}
With controlled mobility	371.12 ± 93.3 s
Without controlled mobility	unsuccessful

C. Experimental result

We limited experiments to 10 min, on the assumption that if the Roomba cannot reach the target node in that time, it never will. Table V shows the results of the experiment. When the Roomba moves according its default algorithm, it cannot approach the target node at all. Using the proposed method, the Roomba reaches target nodes in about 6 min. In a real environment, the effects of shadowing or fading of radio wave are considerable. These effects dynamical change RSSI and therefore, lead the mobile sink into wrong directions. Nevertheless, the mechanism to detect the increase and decrease of the EMA of RSSI allows Roomba to revise the direction of its movement and to approach the target node.

In our simulation, because we adopt the random way point model as the default mobility of the mobile sink, the mobile sink without controlled mobility can reach the target node. However, because the Roomba moves to clean when it is not controlled, it does not necessarily reach the target node.

In terms of overhead of our method, all nodes need to exchange control messages in order to maintain the gradient field at least until a mobile sink reach the target node. However, when monitoring application is assumed and some representative nodes collect sensing information by using a gradient-based routing, additional overhead to control a mobile sink is negligible. In the rescue application, the most important demand is the probability of the guidance of a mobile sink to the target position rather than the overhead of communications.

VI. CONCLUSION AND FUTURE WORK

We proposed a method of controlled mobility using a gradient field. The simulation showed that the proposed method saves the time required for a mobile sink to approach a target node and receive a control message from the target node. We implemented the proposed method on a Roomba in consideration of radio fluctuation in a real environment. We demonstrated that mobile sink mobility can be controlled using a gradient field. The advantage of our proposed method is that any nodes need not to know their own location through the use of devices such as GPS receivers. Moreover, our proposed method can work in indoor environments.

In this paper, we only show the situation that a mobile sink approaches toward the target node gathering sensing data. Our proposed method, however, can be applied to many scenarios; for a patrolling robot or a rescue robot, our method can guide them to the person who requests help. In addition, when a gradient field is constructed according to residual energy, load balancing can be realized.

In future work, we plan to study the case multiple networks exist. In this paper, we use only one network and one target node. Therefore, we need to evaluate whether a mobile sink can approach all the target nodes by rotation when multiple target nodes exist. In this study, we assumed that only one mobile sink exists. We therefore need to consider the case of multiple mobile sinks.

REFERENCES

[1] A. Kansal, M. Rahimi, D. Estrin, W. Kaiser, G. Pottie, and M. Srivastava, "Controlled mobility for sustainable wireless sensor networks," in *Proceedings of the First Annual IEEE Communications Society Conference on Sensor and Ad Hoc Communications and Networks (SECON)*, Oct. 2004, pp. 1–6.

[2] S. Basagni, A. Carosi, E. Melachrinoudis, C. Petrioli, and Z. Wang, "Controlled sink mobility for prolonging wireless sensor networks lifetime," *Wireless Networks*, vol. 14, no. 6, Dec. 2008, pp. 831–858.

[3] R. Sugihara and R. K. Gupta, "Improving the data delivery latency in sensor networks with controlled mobility," in *Proceedings of the 4th International Conference on Distributed Computing in Sensor Systems (DCOSS)*, Jun. 2008, pp. 386–399.

[4] R. Sugihara and R. Gupta, "Optimizing energy-latency trade-off in sensor networks with controlled mobility," in *Proceedings of the 28th Annual Joint Conference on the IEEE Computer Communications (INFOCOM)*, Apr. 2009, pp. 2566–2570.

[5] F. Mourad, H. Chehade, H. Snoussi, F. Yalaoui, L. Amodeo, and C. Richard, "Controlled mobility sensor networks for target tracking using ant colony optimization," *IEEE Transactions on Mobile Computing (TMC)*, vol. 11, no. 8, Dec. 2012, pp. 1261–1273.

[6] R. Falcon, H. Liu, A. Nayak, and I. Stojmenovic, "Controlled straight mobility and energy-aware routing in robotic wireless sensor networks," in *Proceedings of the 2012 IEEE 8th International Conference on Distributed Computing in Sensor Systems (DCOSS)*, May 2012, pp. 150–157.

[7] J. Hao, Z. Yao, and B. Zhang, "A gradient-based multi-path routing protocol for low duty-cycled wireless sensor networks," in *Proceedings of the 2012 IEEE International Conference on Communications (ICC)*, Jun. 2012, pp. 233–237.

[8] R. Nagel and S. Eichler, "Efficient and realistic mobility and channel modeling for vanet scenarios using omnet++ and inet-framework," in *Proceedings of the 1st international conference on Simulation tools and techniques for communications, networks and systems & workshops*. ICST, Mar. 2008, pp. 89:1–89:8.

[9] A. Varga, "Omnet++," in *Modeling and Tools for Network Simulation*. Springer, 2010, pp. 35–59.

[10] C. Bettstetter, G. Resta, and P. Santi, "The node distribution of the random waypoint mobility model for wireless ad hoc networks," *IEEE Transactions on Mobile Computing (TMC)*, vol. 2, no. 3, Jul. 2003, pp. 257–269.

[11] "iRobot Roomba® 500 Open Interface (OI) Specification." [Online]. Available: http://www.robotikasklubs.lv/read_write/file/Piemers/iRobot_Roomba_500_Open_Interface_Spec.pdf [retrieved: June, 2013].

[12] "Crossbow MPR-MIB Users Manual." [Online]. Available: http://bullseye.xbow.com:81/Support/Support_pdf_files/MPR-MIB_Series_Users_Manual.pdf [retrieved: June, 2013].

 ISBN: 978-1-61208-296-7

A 169 MHz Wireless M-BUS Based Advanced Meter Infrastructure for Smart Metering

M. Alonso-Arce, P. Bustamante, G. Solas and J. Añorga
Department of Electronics and Communication
CEIT and Tecnun (University of Navarra)
Donostia-San Sebastián, Spain
maarce@ceit.es, pbustamante@tecnun.es, gsolas@tecnun.es, jbenito@ceit.es

***Abstract*— Nowadays, the use of wireless technologies provides a great benefit to the society. One particular example is the shift towards smarter, more advanced meters. While this trend first started off with the electricity meter market, the adoption of smart meters within the gas and water meter market is also gaining momentum. This article presents an architecture, design and implementation of an advanced meter infrastructure based on 169 MHz, which is suitable for critical deployments as basements, courtyards, etc. The article includes the design and implementation of a wireless sensor node and a gateway for Gas Smart Metering, which can coordinate the network and connect the metering to the "Cloud". This way, the acquired data could be used to improve the measurement systems avoiding human errors.**

Keywords*— *Wireless Sensor Networks, Advanced Metering Infraestrucutre (AMI), wireless M-Bus, Cloud Computing service

I. INTRODUCTION

The cost reduction for the wireless technology has raised the number of applications, in which the remote monitoring and metering are included. This has caused many companies to dedicate their research to the development of wireless sensor platforms, known as *Smart Meters*, aimed at this purpose.

The Smart Meters are defined as advanced energy, water, or gas meters which measure the energy consumption of a consumer and provide added information to the energy provider companies. Smart meters can read real-time energy consumption information including voltage, phase angle and frequency values and securely communicate that data. The smart meters are part of a bigger system where, in addition to the smart meters a bidirectional communication infrastructure and the remote control of the devices are included.

The meter network platform we are proposing is based on the 169 MHz frequency band. This band has been reserved in Europe for meter reading purposes. Along the paper a standard platform and its gateway to the Internet is proposed, developed and tested. The platform is able to read meters or sensors remotely through typical communication protocols such M-Bus, Modbus, I2C, SPI, UART, etc., and also it has been optimized for low cost and low energy consumption.

The meter network is made up of different wireless sensor nodes linked to a master node which is connected to Internet via the mobile network. The sensor nodes are placed near the meters or sensors located in places where a robust communication system is needed, due to the attenuation. Examples of these kinds of places are basements, courtyards, rooftops, etc.

The rest of this paper is structured as follows: in Section II, details are provided concerning the related work including the identification of the main commercial platforms and its characteristics. In Section III, the system architecture is described. Section IV explains the implementation of the different parts of the Advanced Meter Infrastructure. Section V shows both the test done to the system and the results obtained. And finally, conclusions and future work are drawn in Section VI.

II. RELATED WORK

Currently, the commercial available WSN platforms in use are very similar in terms of their hardware architecture, which are based either on the 16-bit MSP430 microcontroller or the 8-bit ATmega128 MCU, in combination with a network architecture based on IEEE 802.15.4 standard. Examples of this type of platforms are TelosB, MicaZ and IRIS. The popularity of the use of the MSP430 and ATmega128 processors is due to several reasons, such as the ultra low power energy consumption, the community support, the open source compilers based on GNU-GCC, and the different embedded operating systems support, such as TinyOS or ContikiOS [1][2]. The limitations of the platforms for the IEEE 802.15.4 based networks have led us to develop a new hardware platform for the 169 MHz frequency band.

Other meter infrastructures are presented in the literature. These meter infrastructures are based on the platforms mentioned above. For example in [3] an AMI is created in the city of Goteborg, based on the ZigBee technology using the 2.4 GHz band. Another example is ZAMI [4], which proposes a ZigBee based infrastructure for automatic meter data collection and energy auditing and management.

In relation with the 169 MHz frequency band, a variety of protocols compete to extend the use of their proprietary approaches. Protocols such as ZigBee [5], EnOcean Routing Protocol [6], KNX-RF [7] and wireless M-Bus are the major

 ISBN: 978-1-61208-296-7

contenders. Currently, the wireless M-Bus has a special importance, as it has been selected by the metering industry, which is organized in associations (i.e. the Open Metering Group).

Wireless M-Bus is based on the Metering bus (M-Bus) which is specialized in transmitting data from the different meters. The wireless M-Bus is specified by the European Norm 13757 [8] and covers the communications from the OSI model hardware layer to the application layer. Although wireless M-Bus does not specify the network organization, leaving it open, gives all the necessary tools to implement it. Wireless companies have developed modules which are in compliance with this norm, having most of them a preprogrammed embedded stack. For this reason, these modules are not suitable for research purposes.

III. System Architecture

The system architecture addresses buildings in a neighborhood, creating a 169 MHz wireless network, in which the meters and the sensors are the network sensor nodes. The communication between each node and the data mining services is carried out via Internet. The interconnection is provided by a special node, the gateway Node. For the connection to the WAN, the mobile communication networks have been selected, using a GSM/GPRS [9][10] module. Due to the fact that some sensor nodes can be out of range with the gateway node, all of the nodes perform message routing.

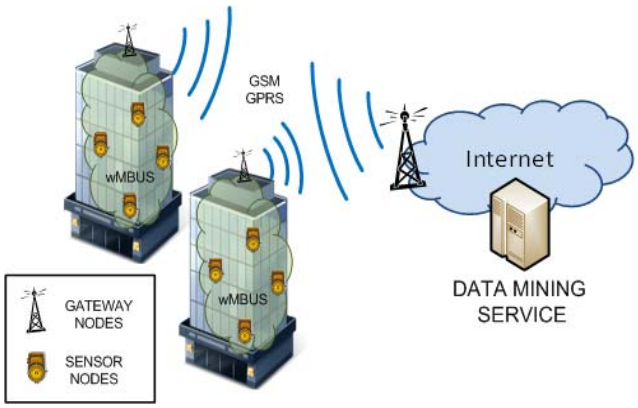


Fig. 1 Advanced Meter Infrastructure (AMI), based on wireless M-BUS networks

Fig. 1 presents the most important components of the system, where each sensor node is able to report data values and respond to external request in an autonomous and independent way. These nodes require small computational power but memory can be an important issue, depending on the time interval the measured values have to be stored.

Usually the sensor nodes are run by batteries; although in some cases they could be run by the electric grid. In the case of the gateway node, it requires significantly higher energy resources for collecting and storing together the data received from the sensor nodes, and to be sent to the data mining services. This node is also responsible for sending the software updates, the configuration commands and test commands to the nodes. As it has been mentioned, this node is connected to the Internet via a GSM/GPRS module. The current consumption of these modules could reach peaks of 1A. This is the reason why the gateway node is powered by the electric grid.

IV. Implementation

A. Hardware implementation of Sensor Node

The objective of developing a new wireless M-Bus platform was to have a base layout design that can be easily controlled and modified. It could also be used in various applications with different interfaces for meters and sensors. Fig. 2 shows a scheme of the proposed Sensor Node.

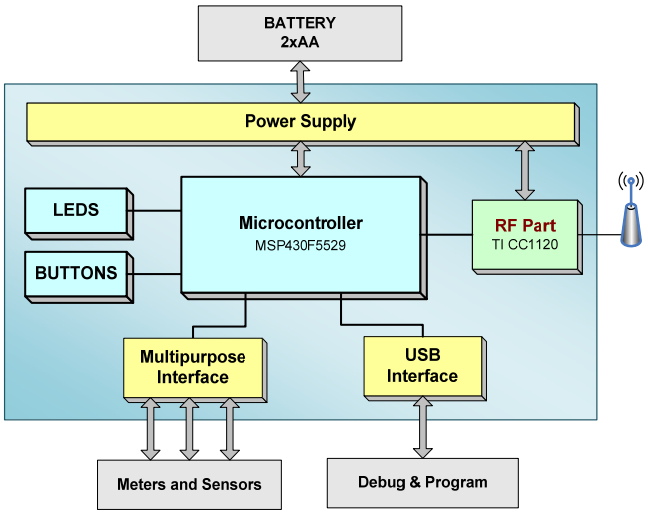


Fig. 2 Sensor's Node hardware scheme

The hardware architecture of the wireless device has been divided in different modules: the microprocessor module, the multipurpose sensor interface, the 169MHz RF module and the Power supply module.

The microcontroller module is the core of the system. It receives the sensor information through the multipurpose interface and also controls the battery level and the radio communications. For this module we chose a MSP430 based microcontroller used in the main commercial platforms such as TelosB, TinyNode, Tmote Sky, etc [11]. The selected one is the MSP430F5529 microcontroller with 8KB+2KB RAM, 128KB Flash and with integrated USB and PHY supporting USB 2.0, four 16-bit timers, a high-performance 12-bit analog-to-digital converter (ADC), two universal serial communication interfaces (USCI), hardware multiplier, DMA, real-time clock module with alarm capabilities, and 47 utilizable I/O pins. The MSP430F5529 is a 16-bit RISC microcontroller equipped with different low power modes between Off Mode (RAM Retention) and Standby mode (1.9μA from 1.8V to 3.6V).

 ISBN: 978-1-61208-296-7

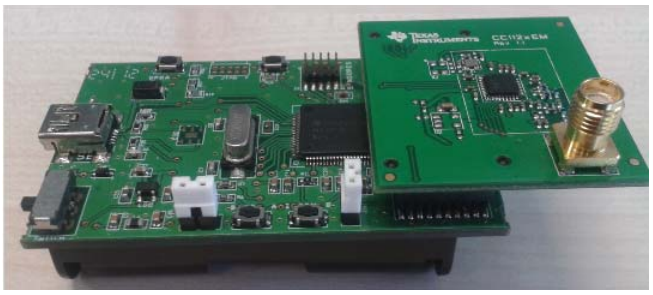


Fig. 3 Picture of the developed sensor node

For the 169 MHz frequency band, the CC1120 chip from Texas Instruments was chosen. This RF module was selected due to the fact that it offers the possibility of changing parameters related to the communication protocol, a feature not provided by the commercial ones.

The Printed Circuit Board (PCB) was designed addressing signal integrity issues for digital circuits such as loop inductance, crosstalk, and Power Distribution Network (PDN) issues. A FR4 2-layer PCB was used with all the components in the top layer. For crosstalk reduction, all empty routing space was filled with cooper and connected to ground plane. In the case of the PDN decoupling capacitors are placed near the IC's power pins minimizing the effect they have on the rest of the circuit. Traces were kept as small as possible and splaying corners are used, to avoid transmission line reflections. In order to have the best performance against the EMC the smallest component packages were used to ensure low equivalent Series Inductance (ESL). The result of the developed sensor node is shown in Fig. 3, where the dimensions are 80mm x 40mm.

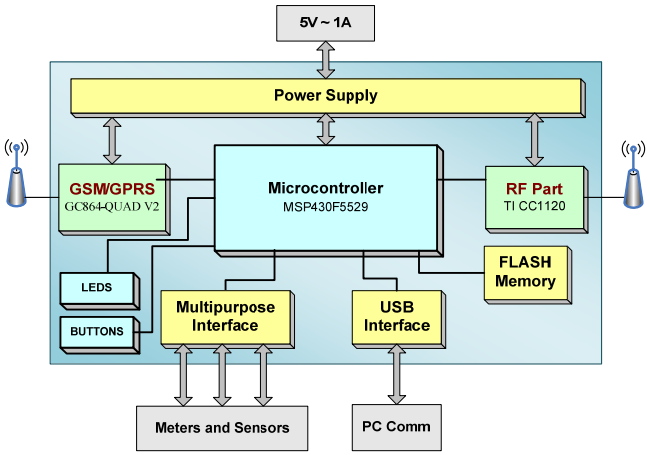


Fig. 4 Gateway's hardware scheme

B. Hardware implementation of the Gateway Node

The gateway node is designed as an extension of the previous node with added modules and features. In this case, a GSM/GPRS module and a FLASH memory are added to the hardware architecture, and the USB interface is used to have a direct connection to the PC. It is used as a command line for debugging and programming. Another remarkable change is the power source which is obtained from the electric grid. A graphical scheme of the gateway node is shown in Fig. 4.

The PCB was designed similarly to the one of the sensor node. The result of the gateway node implementation is shown in Fig. 5, where the dimensions are 85mm x 54 mm.

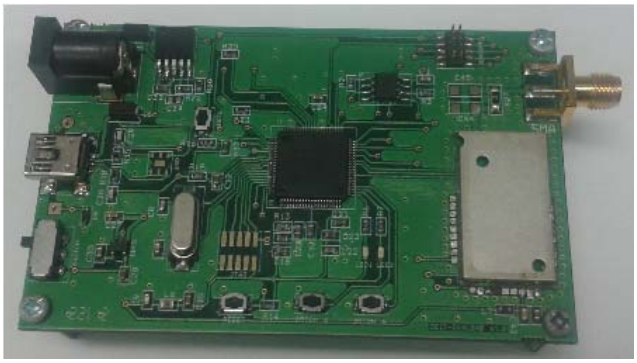


Fig. 5 Picture of the developed wireless M-BUS Gateway

C. Software implementation

The developed software is a key point of the system and has a modular scheme. It also created from scratch and built with the open msp430-gcc compiler. The design allows the different modules to be independent from the hardware as long as the same standards and definitions are used. Fig. 7 shows the software architecture and the interactions between its components. The major components are:

- **Sensor node software:** reports information about the sensors and responds to requests.
- **Gateway node software:** includes the following functions: request data, aggregate nodes to the network, manages communications and provides information about all the nodes to the "cloud".
- **Data mining service:** exploits the data and introduces them into the database.
- **Web interface:** displays the data and allows the user to send commands to the network or to an individual node.

The sensor node software implements a wireless M-Bus stack besides a power saving algorithm, designed to manage all the power modes of the modules. It puts the modules in the best mode, considering the tradeoff between their functionality and the minimum power consumption.

```
<device>
  <id>01:02:0A:0B:0C:0D:01:01</id>
  <measure>
    <var>
      <varid>1</varid>
      <value>2</value>
      <time>2013-04-16 12:20:53</time>
    </var>
  </measure>
</device>
```

Fig. 6 XML file description

 ISBN: 978-1-61208-296-7

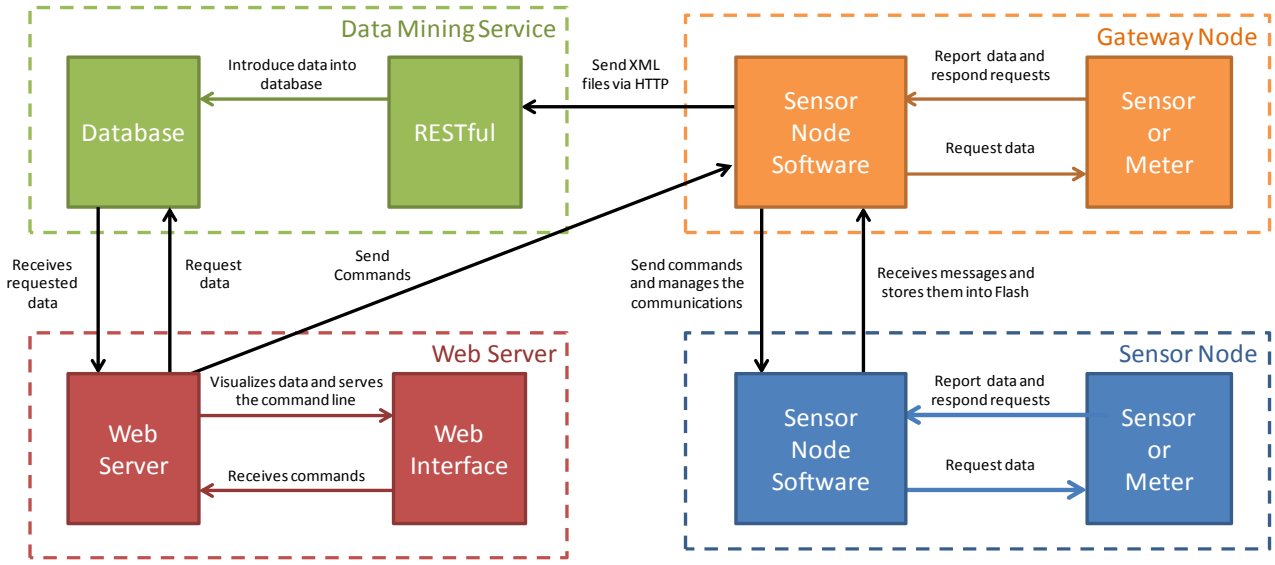


Fig. 7 Software Architecture

The embedded software running in the gateway node also implements the wireless M-Bus stack, but in this case it is not worth implementing the power saving algorithm. Furthermore, this software implements the gateway functionality, which translates the data between the wireless M-Bus network and the Internet. However, another protocol translation occurs prior to this one. Data obtained from the wireless M-Bus network is translated into a custom XML, and is sent to the data mining service by the HTTP protocol over TCP/IP via the GSM/GPRS module. In addition, the gateway node has a USB interface with a console for configuration. An example of a piece of code from a generated XML file is shown in Fig. 6.

The data mining service is composed by a RESTful web service API, a MySQL database and a web server service. The RESTful service receives the XML file generated and sent by the gateway nodes and introduces it into the database. The web server service is used as a web interface that handles the data visualization and is able to configure some parameters of the nodes and also to send commands to the gateway.

V. Tests and Results

In this section, the most relevant results and tests have been evaluated. First, a power consumption study of the main operations of the sensor node is shown. Then the evaluation of the gateway communications front-end is exposed. Finally, the whole system with Internet connection is evaluated and analyzed.

A. Sensor Node

The energy consumption of the developed sensor node has been studied by analyzing the voltage drop across a 10Ω resistor placed in series with the power source. Then the current consumption is obtained by applying the Ohm's law.

Fig. 8 shows the current consumption profile obtained for different typical operations.

- Interval 1 and 5 shows the device in the best power down mode, where the average current consumption is 0.223mA.
- Interval 2 corresponds to the microcontroller in active mode and the transceiver in idle mode, where the total current consumption average is 2.529mA.
- Interval 3 relates to the receive operation of the transceiver in the channel "1a", defined in the wireless M-Bus standard. In this interval the device is waiting for a packet and receives it, in which case the current consumption average is 25.002mA.
- Interval 4 presents a transmit operation of a wireless M-Bus packet in channel "1a", composed by 23 bytes without taking into account the preamble. In this case the current consumption average is 40.075mA during 45.5ms.

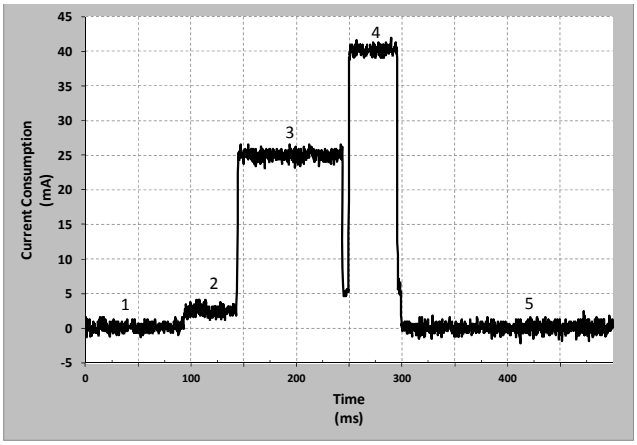


Fig. 8 Node´s Energy consumption

B. Gateway Node

The gateway node was built and tested similar to the sensor node. Upon the electrical test was done, the connection to the PC through the USB was tested, creating a command line in the USB and configuring the USB line as a serial CDC emulation. The GSM/GPRS module was tested by sending AT commands via the command line and observing the responses, sending some text messages to some phone devices.

C. Advanced Meter Infraestrucutre

The system was prepared by setting-up a wireless meter network with 2 sensor nodes and 1 gateway node. As the sensor nodes do not have connected any sensor or meter, the sensor nodes were configured to send periodical messages every 2 minutes with the "On time" value to the gateway node. The data mining service and the web interface was configured as it is presented in the implementation.

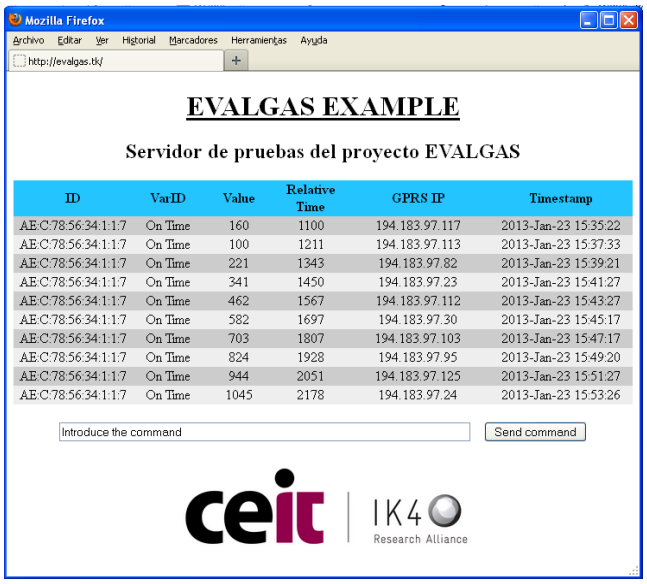


Fig. 9 Web interface screenshot

The infrastructure was tested analyzing the results obtained in the web interface, which shows the "On time" values configured to be sent on sensor nodes. A screenshot of the web interface is shown in Fig. 9. The developed wireless M-Bus stack was checked by requesting data of a sensor node. The results obtained validate the network functionality and the suitability of the wireless M-Bus for the AMI.

VI. Conclusions and Future Work

This paper proposes an architecture, describes the implementation of the required hardware and presents the design and the implementation of the software for an Advanced Meter Infrastructure. We successfully build a small low power sensor node platform and a gateway node platform.

Wireless M-Bus demonstrated to be a good solution to implement the communications for the Advanced Meter Infrastructure. The software architecture has been validated and allows bidirectional communication between the web interface and the sensor nodes.

In addition, a power consumption analysis has been performed and the obtained value will be used to determine the periods of transmission for a specific lifetime.

Currently we are working on improving and enlarging the knowledge of the meter infrastructure in several ways:

- Defining a new generic XML file.
- Studying the transmissions from a physics point of view.
- Introducing the *Internet of Things* to the system, allowing TCP/IP communications.
- Release the Hardware and Software as an open Hardware Platform for Metering purposes.

Acknowledgment

This work was partially supported by EVALGAS project within the framework of ETORGAI projects, founded by the Basque Government.

References

[1] M. Kuorilehto, M. Kohvakka, J. Suhonen, P. Hämäläinen, M. Hännikäinen, and T.D. Hamalainen, "Ultra-Low Energy Wireless Sensor Networks in Practice", John Wiley & Sons, 2008, Part II, pp. 29-49.

[2] M. Johnson, M. Healy, P. van de Ven, M.J. Hayes, J. Nelson, T. Newe, and E. Lewis, "A Comparative Review of Wireless Sensor Network Mote Technologies", IEEE Sensors Conference, 2009, pp. 1439-1442.

[3] Thomas Arnewid, "AMI in the city of Goteborg", Metering International, Issue 4, 2007, pp 76-79.

[4] Hoi Yan Tung, Kim Fung Tsang, and Ka Lun Lam, "ZigBee sensor network for Advanced Metering infrastructure", International Conference on Consumer Electronics, Digest of Technical Papers, 2010, pp. 95-96.

[5] ZigBee Alliance, "ZigBee: Wireless Control That Simply Works", ZigBee White Papers, April 2007.

[6] EnOcean GmbH, "Routing Concept from EnOcean", EnOcena Green Smart Wireless White Paper, February 2006.

[7] KNX and EnOcean, "KNX and EnOcean White Paper", enocean alliance, 2008.

[8] "Communication systems for meters and remote reading of meters", EN 13757, 2011.

[9] A.H. Primicanta, M.Y. Nayan, and M. Awan, "Hybrid Automatic Meter Reading System", Computer Technology and Development International Conference, 2009. Vol.2, pp. 264-267.

[10] F. Ding, G. Song, J. Li, and A. Song, "A ZigBee Based Mesh Network for Home Control System", Education Technology and Training and International Workshop on Geoscience and Remote Sensing, 2008. Vol.1, pp. 774-748.

[11] H.Y. Tung, K.F. Tsang, and K.L. Lam, "ZigBee sensor network for Advanced Metering infrastructure", International Conference on Consumer Electronics, Digest of Technical Papers, 2010, pp. 95-96.

 ISBN: 978-1-61208-296-7

Transceiver-power Control for 802.15.4a UWB-IR Ranging in the Presence of Multipath Propagation

Tingcong Ye
Tyndall National Institute
University College Ireland, UCC,
Cork, Ireland
Email: tingcong.ye@tyndall.ie

Michael Walsh
Tyndall National Institute
Cork, Ireland
Email: michael.walsh@tyndall.ie

Brendan O'Flynn
Tyndall National Institute
Cork, Ireland
Email: brendan.oflynn@tyndall.ie

Cian O'Mathuna, $Fellow, IEEE$
Tyndall National Institute
Cork, Ireland
Email: cian.omathuna@tyndall.ie

***Abstract*—Sensor networks can benefit from location-awareness, since it allows information gathered by sensors to be tied to their physical locations. The emerging IEEE 802.15.4a Ultra Wideband Impulse Radio (UWB-IR) transmission is a promising technology for ranging and location-aware sensor networks, due to its fine time resolution and power efficiency. However, the presence of multipath propagation presents a significant challenge in terms of ranging, as they can result in biased distance estimates. In this paper, we present multipath effects mitigation methods using bilateral transceiver-power control algorithms, which are capable of: (i) optimizing the signal to noise ratio (SNR) of leading path signal and maintaining the connectivity; and (ii) cooperating with a practical symmetric double-sided two-way ranging protocol. Relevant aspects of power control are discussed using an automatic control framework. We evaluate the resulting performance and compare with existing non-power control techniques. Experimental results show that the proposed multipath effects mitigation approach is more robust against the non-power control based ranging errors.**

***Keywords*—802.15.4a UWB-IR, Transceiver-power Control, Ranging, Multipath Propagation.**

I. INTRODUCTION

Location-awareness becomes an essential aspect of wireless sensor networks (WSNs) and will enable a wide variety of applications, in both the commercial and the military sectors. Ultra Wideband Impulse Radio (UWB-IR) transmission provides robust signaling, as well as high resolution ranging capabilities [1]. Therefore, UWB represents a promising technology for ranging and localization applications through time-of-arrival (TOA) technique [2], [3]. IEEE has recognized the need to standardize UWB technology for use in personal area networks (PANs) and has established the IEEE 802.15.4a standard specifying a new UWB physical layer for WSNs [4]. In practical scenarios, however, a number of challenges remain before UWB ranging and communication can be deployed. These mainly include uncommon time reference, clock drift, low sampling rate, multipath propagation and non-line-of-sight (NLOS) propagation. The use of practical Symmetric Double-Sided Two-Way (SDS-TW) ranging protocol [4] and Energy Detector based TOA estimation methods [5]-[7] can significantly solve the uncommon time reference, clock drift and low sampling rate issues. In real-world environment, the multipath propagation issue is especially critical for 802.15.4a ranging system, since uncertain multipath effects (like reflection, scattering and diffraction) introduce noise sources in the TOA estimation, thus seriously affecting the Signal-to-Noise (SNR) of the leading path signal and ranging performance. Indoor environments have a high occurrence of multipath propagation situations. It is therefore critical to understand the impact of multipath propagation on 802.15.4a ranging systems and to develop techniques that mitigate their effects.

Super-resolution TOA technique is proposed to estimate the channel multipaths TOA and amplitude using MUSIC [8]. A pattern to identify closely-spaced direct path and ground reflection is given in [9]. The main drawback of existing multipath mitigation techniques is difficulty in determining the direct path and the large number of multipath components especially when paths overlapping. The Cramer-Rao lower bound of TOA shows that SNR is, a more accurate indication of the link status and ranging accuracy [10]. This paper considers a channel quality improvement approach. In particular, we propose the use of power control to optimize the SNR and thus performing multipath mitigation. Power control algorithms for wireless communication systems have been investigated and designed to satisfy the QOS requirements [11]. The IEEE 802.15.4 (Zigbee) based mobile tracking system has used the power control methods to solve the location error [12]. In this work, we firstly performed an extensive indoor measurement campaign with FCC-compliant 802.15.4a UWB radios to quantify the effect of multipath propagation. From these channel impulse responses, we extract features that are representative of the multipath propagation conditions. We then develop multipath effect mitigation algorithms based on bilateral transceiver-power control protocol. Performance comparison is made with the non-power control scheme.

The rest of the paper is organized as follows. Section II introduces a practical ranging method and SNR estimation. Section III analyzes the features of multipath propagation. Section IV describes bilateral power control algorithms. Numerical

 ISBN: 978-1-61208-296-7

performance results are provided in Section V, and we draw our conclusions in Section VI.

II. IEEE 802.15.4A RANGING AND SNR CALCULATION

To solve the uncommon time reference and clock drift issues, 802.15.4a ranging system [14] employs a Symmetric Double-Sided Two-Way (SDS-TW) ranging protocol [4] and its architecture is illustrated in Fig.1 (a). It requires both transceivers to record time-of-transmitting (TOT) values and TOA values to measure round-trip times (RTT) and the time-of- flight (TOF). In practical implementation, the TOF calculation can be expressed as:

$$TOF = \frac{RTT_L - D_F + RTT_F - D_L}{4} \quad (1)$$

Where, $RTT_L = TOA2 - TOT1$ is the round-trip time of leader, $D_F = TOT2 - TOA1$ is the response delay of follower, $RTT_F = TOA3 - TOT2$ is the round-trip time of follower, and $D_L = TOT3 - TOA2$ is the response delay of leader. And the distance is:

$$d = \dot{c} \times TOF \quad (2)$$

Where, d is the antenna-to-antenna distance and $\dot{c}$ is the speed of electromagnetic waves.

A. Performance Limits of TOA-based Ranging

Ideally, TOA estimate is given by the time instant corresponding to the maximum absolute peak at the output of the marched filter over the observation interval. The performance of the estimator achieves the Cramer-Rao Lower Bound (CRLB) for signal SNR in AWGN channel as described in [10]. The mean square error of any unbiased estimate $\hat{\tau}$ of the true TOA τ can be lower bounded by:

$$V(\xi) \geq \frac{1}{2\sqrt{2}\pi B\sqrt{SNR}} \quad (3)$$

Where, the $V(\xi)$ is the variance of TOA error, $\xi = \widehat{\tau} - \tau$ is the TOA error. B is the effective signal bandwidth. It shows that SNR is a more accurate indication of the ranging performance.

B. SNR Calculation of 802.15.4a Impulse Signal

The SNR, computed by the receiver, is equal to the difference between the power of received signal and the noise power. The received signal power depends on the transmission power, the distance between transmitter and receiver, the environment (walls, obstacles will reflect and distort the signal creating a multipath effect) and interference from other sources. There are many SNR models have been proposed, one is the most correct SNR model [13] is:

$$SNR = (\sum_{i=1}^{N}(SNR_i)^{-1})^{-1} \quad (4)$$

Where, $SNR_i = E_r/E_i$, E_r is the desired received signal power, E_i is related to i^{th} noise sources including multipath components, multiple-user interference and thermal noise. It is worth noting that due to multipath effects, a Gaussian random path loss variable $N(0, \sigma)$ exists in the path loss model with a zero mean and σ standard deviation.

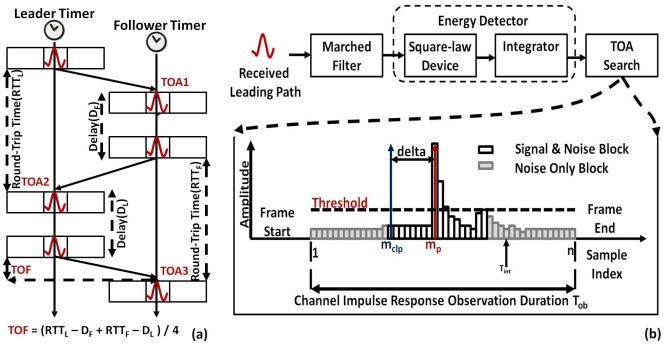


Fig. 1: (a) SDS-TW protocol and (b) Energy Detector based TOA estimator.

The SNR of the received leading path signal is measured in the leading path channel impulse response in this wrok, see Fig.1 (b). The received leading path signal is first passed through a marched filter (MF). The observed signal forms the input to the Energy Detector (ED), whose output is sampled at every T_{int} seconds, T_{int} is the tick time units. T_{ob} is the Channel Impulse Response (CIR) observation duration. $n = T_{ob}/T_{int}$. The system searches the index with peak power m_p in the CIR. The received signal power is with two-sided power spectral density in the CIR and the m_p is the central axis. According to signal sampling duration T_s, the effective signal duration in the CIR should be $2\kappa T_{int}$,$\kappa = T_s/T_{int}$. The total signal power is then measured by $P = \sum_{i=m_p-2\kappa}^{m_p+2\kappa} P_i$. P_i is the power of i^{th} index. The average power of all indices of the effective signal is $P_a = P/4\kappa$. The deviation is then calculated as the noise part $P_n = (\sum_{i=m_p-2\kappa}^{m_p+2\kappa}(p_i - P_a))/4\kappa$. Hence, the effective signal part is calculated as $P_r = (\sum_{i=m_{clp}}^{m_{clp}+2\kappa} p_i) - 2\kappa P_n$. m_{clp} is the time stamp window index, as indicated by first byte of channel response read. The SNR of received leading path is:

$$SNR = 10\log(P_r/P_n) \quad (5)$$

III. MULTIPATH EFFECTS OBSERVATION

The aim of this experimental effort is to build a database containing a variety of propagation conditions (without multipath and with multipath). The measurements were made using two the world's first IEEE 802.15.4a standard compliant UWB-IR transceivers that were developed by Decawave Company [14]. Both transceivers (called leader and follower respectively) set up a typical ranging system to measure the antenna-to-antenna distance and record the ranging measurements. The 802.15.4a UWB-IR ranging parameters are listed in Table I. To compare the features of non-multipath and multipath propagation, ranging measurements were implemented in an anechoic chamber (AC) (without multipath propagation) and in an indoor office (OFC) (with multipath propagation), using a distance of 0.7m in LOS condition, see Fig.2 (a) and (b). Various transmission ranges are selected to capture a variety of operating conditions from 2.5m to 30m, see Fig.2 (c). The ranging error is shown using the empirical cumulative distribution function (CDF).

The Channel Impulse Response observation duration is 992

 ISBN: 978-1-61208-296-7

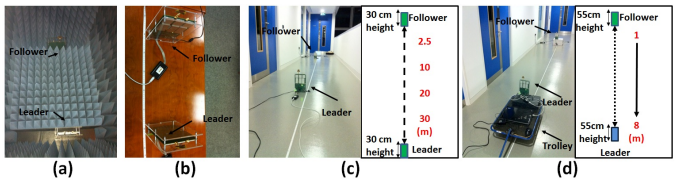


Fig. 2: Ranging Experimental Setup: (a) Anechoic Chamber; (b) Office (c) Hallway; (d) Hallway (mobile).

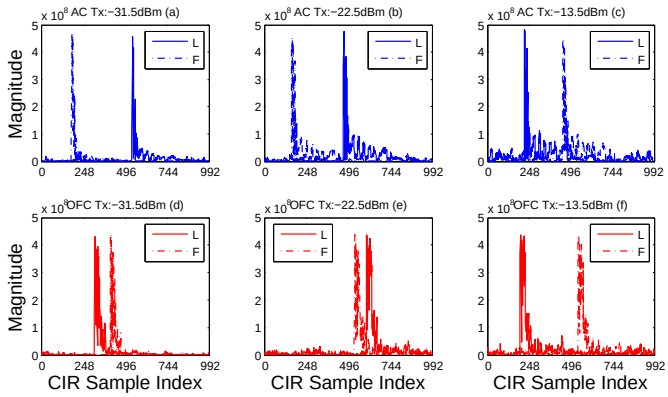


Fig. 3: Channel Impulse Response captured at anechoic chamber (a), (b), (c) and indoor office (d), (e), (f).

indices. The LSB of a time value represents 1/128 of a chip time at the mandatory chipping rate of 499.2 MHz. Indices of signal are samples 998.4 MHz at 500KHz bandwidth, which is 64 times than the tick time units.

A. *The impact of Multipath Effects*

Fig.3 shows some channel impulse responses (CIRs) captured using different signal outputs (-31.5dBm, -22.5dBm and -13.5dBm) in anechoic chamber (AC) and an indoor office (OFC). Without multipath effects in AC, noise becomes more evident at higher transmission powers. Few adjacent peaks with comparable amplitude to the largest peak are exhibited in the CIRs and leading path detection more straightforward. With mulitpath effects in real office (OFC) environment, more

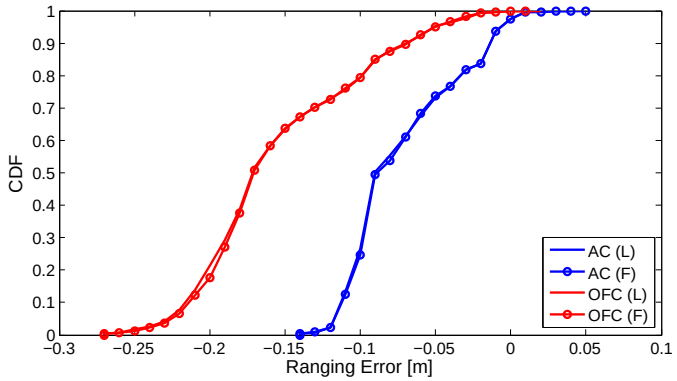


Fig. 4: CDF of the ranging error in AC and OFC.

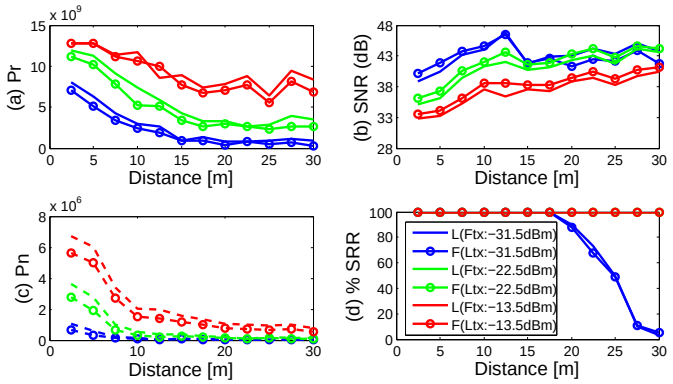


Fig. 5: Parameters measurements in different locations (a) Pr; (b) SNR; (c) Pn and (d) PRR.

TABLE I: IEEE 802.15.4a UWB Signal Parameters

IEEE 802.15.4a Channel index	2
Preamble Length	1024
Pulse Repetition Frequency	16MHz
Data Rate	850Kbits/s
Signal Bandwidth	500 MHz
Center Frequency	4 GHz
Maximum Transmit-power	-13.5dBm

adjacent peaks with comparable amplitude to the largest peak are exhibited in the channel impulse responses (CIR). The adjacent peaks increase the possibility that a reflection is selected as the leading path resulting in the introduction of ranging inaccuracy. Fig.4 shows the CDF of the ranging error. In the anechoic chamber, the ranging error below 15cm/ 10cm occurs 100% and 85% of the 7500 measurements, respectively. In the indoor office, the ranging error below 15cm/10cm occurs 20% and 5% of the 7500 measurements, respectively. Hence, we can say this practical ranging method is reliable and accurate; and the multipath propagation adversely affect the ranging performance.

B. *802.15.4a Multipath Propagation Channel*

In indoor environments, due to various obstructions, the multipath effect is uncertain and random. Fig.5 shows the multipath channel features ranging from 2.5m to 30m, using different signal outputs (-31.5dBm, -22.5dBm and -13.5dBm). At a fixed distance, the higher the transmit-power, the lower the SNR, see Fig.5 (b). With the distance increasing, Received signal power (Pr) decreases almost linearly, see Fig.5 (a). However, the SNR increases with distance increasing before 10m, see Fig.5 (b). After 10m, the SNR obtained using minimum transmit-power (-31.5dBm) decreases. However, the SNR increases when using higher transmit-power (-13.5dBm). The noise power (Pn) decreases and reaches to the noise floor, see Fig.5 (c). It means that the multipath effects become less impact on the UWB signal transmission with distance increasing. The greater the transmission range, the poorer the connectivity and the lower the signal receive rate (SRR), see Fig.5 (d). Hence, there is a tradeoff between SNR and connectivity.

 ISBN: 978-1-61208-296-7

IV. Bilateral Transmitter Power Control Algorithms

A. Power Control Aspects

Being subjective, some relevant aspects of the power control in 802.15.4a UWB ranging system should be considered.

• Ranging algorithm. This is the basic framework for UWB ranging system. Power control loop should be integrated into this framework.

• Power constraints. The transmission powers are considered due to the device limitations and the FCC limits.

• Time delays. Measuring and control signaling take time, resulting in time delays are typically fixed due to standardized signaling protocols.

• Tradeoff management. The one is the most correct BER expression [1] is:

$$BER = \frac{1}{2} erfc(\sqrt{\frac{1}{2} SNR}) \tag{6}$$

Where, the $erfc$ is the error function. The BER is used to manage the tradeoff between receive-power and SNR. A set-point SNR value can be selected and related directly to BER.

B. Power Control Methods

According to the ranging algorithm, see Fig.1(a), a TOF is calculated through the bilateral communication links between the leader and the follower. Hence, both UWB transceivers need to implement the power control loop integrated into the ranging algorithm. Consider the arrangement of power control scheme at one UWB transceiver in Fig.6, the transmit-power level is increased/decreased depending whether the SNR decoded ($\gamma_j(t)$) from the j^{th} remote signal $r_j(t)$ is above or below the target SNR (γ^T), and implemented as:

$$e_j(t) = \gamma_j(t) - \gamma^T \tag{7}$$

$$p_j(t+1) = p_j(t) + \Delta e_j(t) \tag{8}$$

$$\begin{aligned} p_j(t+1) =& min\{P_{max}, p_j(t) + \Delta e_j(t)\} \\ & \| \; max\{P_{min}, p_j(t) + \Delta e_j(t)\} \end{aligned} \tag{9}$$

From equation (7) to (9), the $e_j(t)$ is the power update command, Δ is the minimum interval of the device power settings. P_{max} is maximum transmit-power of the FCC limits and P_{min} is the minimum power of the device limits. The P_{max} and P_{min} are controlled to ensure the inherent hardware saturation limitations are not surpassed. $p_j(t+1)$ is an integrating controller with $C_j\{\Delta e_j(t)\} = \Delta e_j(t)$ in one-slot cycle. This can be denoted as quick power control (QPC). At the same time, the SNR of the leading pulse measured at local mote $\gamma_i(t)$ is sent back to the remote UWB mote for remote power control implementation.

Another power control algorithm is that makes it possible to emulate slower update rates, or to turn off power control by transmitting a series of $e_j(t)$. It can be denoted as slow power control (SPC). In a n-slot cycle ($k = 1...n$), the power update command $e_j(t)$ in equation (10) is computed according to:

$$e_j(t) = \frac{\sum_{k=1}^{n} e_k(t)}{n} \tag{10}$$

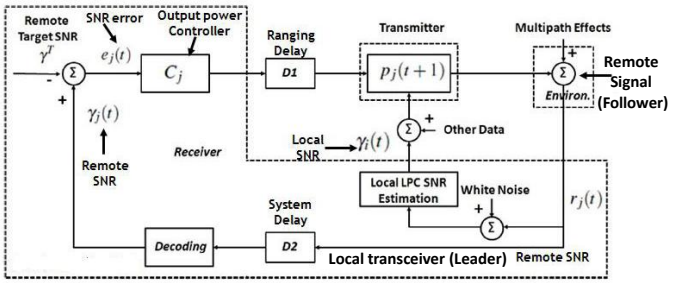


Fig. 6: Block diagram of bilateral transmitter output power control loop at local transceiver.

The decision feedback where the sign of error $e_j(t)$ is fed back resulting in equation (11):

$$s_j(t) = sign(\gamma_j(t) - \gamma^T) \tag{11}$$

An adaptive fixed step size power law (FPC) is considered. The power $p_j(t)$ is increased or decreased by βdBm depending on the sign of error $s_j(t)$ as equation (12):

$$p_j(t+1) = p_j(t) + \{^{-\beta, s_j(t)<0}_{+\beta, s_j(t)>0} \tag{12}$$

In the ranging architecture, see Fig.1 (a), the delay, is set for mitigating the clock drift issue (about 300ms). Hence, cooperating with this ranging algorithm, there is enough time for measuring and control signaling delay ($D1$) in the power control loop. The other delay ($D2$) is the system delay which is also considered into the ranging algorithm.

V. Practical Implementation

The above power control methods and a pre-existing non-power control (NPC) method are critically assessed using a multipath propagation issue scenario. Stationary (from 2.5m to 30m) and mobile (from 1m to 8m with a trolley) ranging experiments are set up inside a real building, see Fig.2 (c) and (d). A target SNR value of 43dB is selected for both UWB transceivers, guaranteeing a BER of $< 3e^{-11}$, verified using equation (6). The Δ is set to be 1.5dBm and the $e_j(t)$ is round to integer. The maximum transmit-power (-13.5dBm) is selected at the beginning for strongest links. A frame is a record of the receive-time of the leading pulse according to the ranging algorithm. Basically, a ranging frame-time is about 0.6s which is equal to 2 delays (300ms) plus 2 TOFs.

A. Stationary power control test

For the purposes of clarity, the response of power control methods at a distance of 17.5m are presented graphically in Fig.7 to Fig.8. The NPC uses maximum transmit-power (-13.5dBm), see Fig.7 (d), and the response shows that system can not reach to the target SNR (43dB), see Fig.8 (d). The QPC method, using equations (7)-(9), updates the signal outputs per frame, see Fig.7 (a), and the instant SNR measured reaches the target SNR (43dB) at the first frame, see Fig.8 (a). The SPC method updates signal output every 10 frames, see Fig.7 (b), and maintains the target SNR after about 15 frames. The FPC method updates with a fixed step size, reaching to the target SNR is slower than the QPC but faster than the SPC, see Fig.7 (c) and Fig.8 (C).

 ISBN: 978-1-61208-296-7

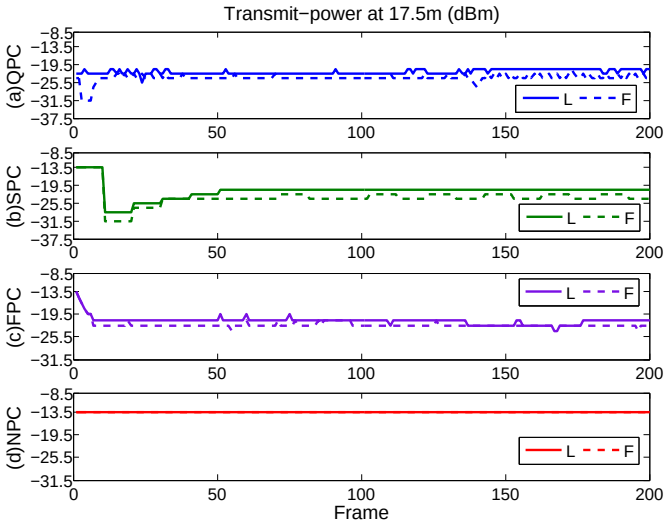


Fig. 7: Transmit-power updating at 17.5m.

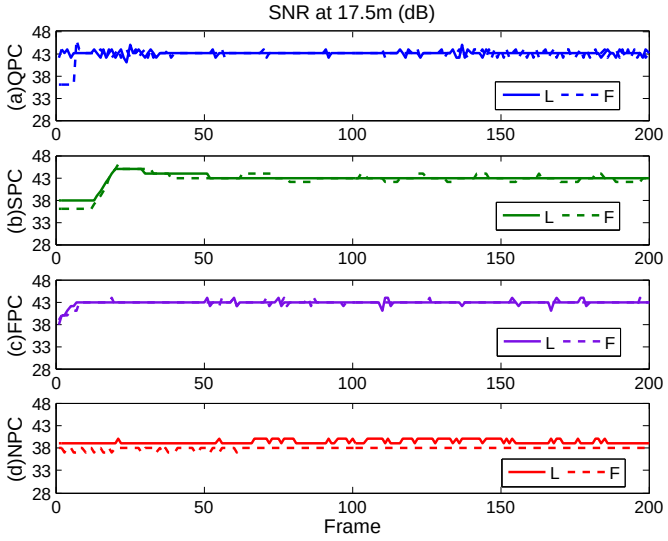


Fig. 8: SNR updating at 17.5m.

Fig.9 shows the CDF of the ranging error measured using different power control methods. Table II summaries the results comparison of different power control algorithms. Using the QPC method, ranging error below 10cm/5cm occurs in more than 91%/60% of the measurements, respectively. The worst case is 18cm. SPC method gets the ranging error that below 10cm/5cm occurs in more than 85%/50% of the measurements, respectively. The worst case is 60cm. FPC method obtains the most accurate ranging estimates, its ranging error below 10cm/5cm occurs in more than 94%/80% of the measurements, respectively. The worst case is 20cm. Without power control, using the NPC method, ranging error below 10cm/5cm occurs in less than 20%/5% of the measurements, respectively. The worst case is greater than 1m.

To measure the transmit-power efficiency for the respective algorithms, the transmit-power efficiency for any one controller configuration is defined as the average transmit-power consumed by two transceivers operating using a particular power control algorithm for the duration of an experiment. For example 100% transmit-power efficiency in this context would imply that the mote is transmitting using its minimum output power setting. Fig.10 plots the percentage transmit-power efficiency for each of the mote control configuration. The utilization of power control methods can get more transmit-power efficiency (up to 99%) than non-power control method (47%). With the distance increasing from 2.5m to 30m, the transmit-power efficiency of the power control methods decreases from 99% to 50%, that because of the controller increases the output power to maintain SNR and connectivity.

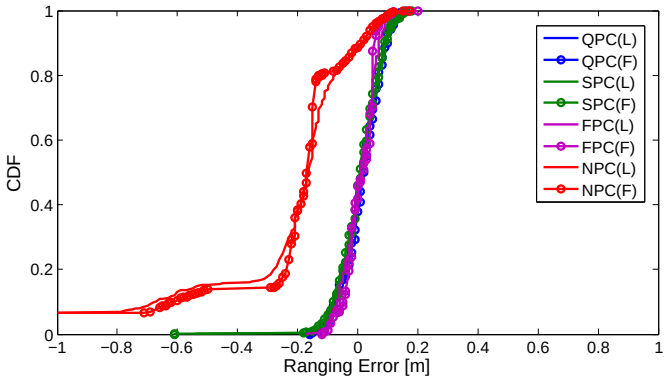


Fig. 9: CDF of the ranging error for stationary test using power control methods from 2.5m to 30m.

TABLE II: Stationary Ranging Results Comparison of QPC, SPC,FPC, NPC

	QPC	SPC	FPC	NPC
Power Update Rate (cycle)	1	10	1	0
Power Update Value (dBm)	$1.5e_j(t)$	$1.5e_j(t)$	$+/-1.5$	Null
Measurements	4800	4800	4800	4800
Accuracy $> 10cm$	>91%	>85%	>94%	<20%
Accuracy $> 5cm$	>60%	>50%	>80%	<12%
Worst Case	18cm	60cm	20cm	0-> $1m$

B. *Mobile power control test*

The mobile ranging test is made to observe the performance of the power control method in the real-world environment with uncertain factors such as the motion of the leader and time-varying wireless channel. Being subjective, the SPC method with lower power updating rate is not powerful for the mobile ranging when compared to the QPC and FPC methods. Thus, in this section, we only consider the NPC, FPC and QPC. The follower (F) is stationary for the duration of the experiment. The leader (L) moves from 1m (from the follower) in a straight line to a distance of 8m with an approximate constant velocity of (20.4 sec/m), see Fig.2 (d).

The QPC (blue line) and FPC (violet line) methods adjust transmit-power according to whether the SNR is greater or less than the target SNR (43dB), see Fig.11 (a). The NPC method keeps the maximum transmit-power (red line). Before the 100^{th} frame point, both UWB transceivers employing QPC

 ISBN: 978-1-61208-296-7

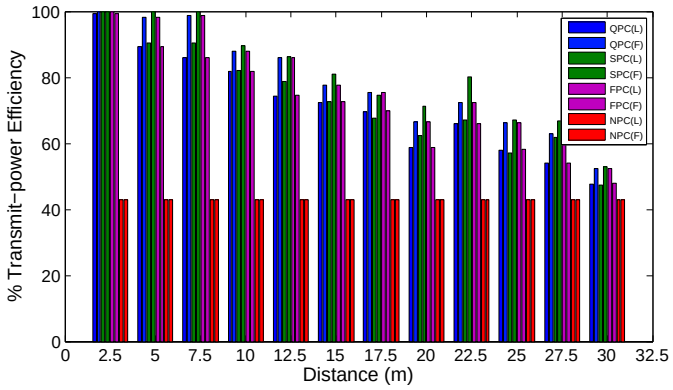


Fig. 10: Transmit-power efficiency from 2.5m to 30m.

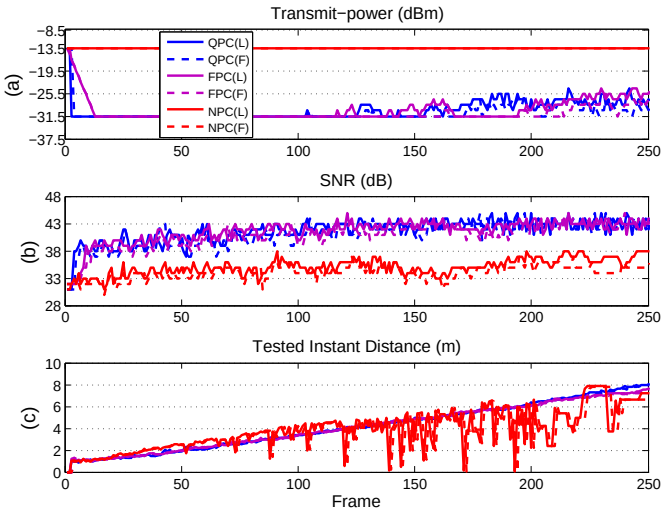


Fig. 11: Transmit-power control in mobile condition.

and FPC use minimum transmit-power (-31.5dBm) and the SNRs are less than 43dB. However, the value of the SNRs increases with distance increasing, see Fig.11 (b). Whereas, the SNRs obtained from the NPC are always less than 38dB, but the value increases with distance increasing from the starting point (32dB). In Fig.11 (c), the leader moves from 1m (starting at about 13 frames point) to the end (8m test-point) with a approximate velocity (34frames/m). When employing power control methods (QPC and FPC), the slopes of the ranging curve of both transceivers approximately meet the velocity. The power controlled ranging channel is more stable than non-power control (NPC) which obtains highly variable ranging estimates (0 $\sim$ 4m difference) during the moving period.

VI. Conclusion

In this paper, we presented the bilateral transmitter output power control algorithm to deal with the effects of multi-path propagation. This technique does not require changing the ranging protocol and time-of-arrival estimation method. To validate this technique in realistic scenarios, we performed an extensive indoor measurement campaign using FCC-compliant IEEE 802.15.4a impulse UWB radios. The bilateral power control method that is capable of dynamically controlling the outputs of the 802.15.4a transceivers to optimize the SNR of leading path signal. We observe that our bilateral power control method can: (1) cooperate seamlessly with a practical symmetric double-sided two-way ranging method; (2) compensate for uncertain multi-path effects and maintain connectivity and (3) significantly improve the ranging performance for multipath environments for both the stationary and mobile case in line-of-sight conditions.

Acknowledgment

This work was supported by Networked Embedded Systems (NEMBES) project, High Education Authority (HEA) Programme for Research in Third Level Institutions (PRTLI)-Cycle 4.

References

[1] M. Z. Win and R. A. Scholtz. “Characterization of ultra-wide bandwidth wireless indoor communications channel: A communication theoretic view”. *IEEE J. Sel. Areas Commun.*,vol. 20, no. 9, pp. 1613-1627, Dec. 2002.

[2] J.-Y. Lee and R. A. Scholtz. “Ranging in a dense multipath environment using an UWB radio link”. *IEEE J. Sel. Areas Commun.*,vol. 20, no. 9, pp. 1677-1683, Dec. 2002.

[3] D. B. Jourdan, D. Dardari and M. Z. Win. “Position error bound for UWB localization in dense cluttered environments”. *IEEE Trans. Aerosp. Electron. Syst.*,vol. 44, no. 2, pp. 613-628, Apr. 2008.

[4] IEEE working group 802.15.4a. Wireless Medium Access Control (MAC) and Physical Layer (PHY) Specifications for Low-Rate Wireless Personal Area Networks (WPANs). Alternate PHYs. NJ: IEEE, 2007.

[5] P. Cheong, A. Rabbachin, J. Montillet, K. Yu and I. Oppermann. “Synchronization, TOA and position estimation for low-complexity LDR UWB devices”. *IEEE Int. Conf. Ultra-Wideband (ICUWB), Zurich, Switzerland*, pp.480-484, Sep. 2005.

[6] A. A. DAmico, U. Mengali, and L. Taponecco. “Energy-based TOA estimation”. *IEEE Trans. Wireless Commun.*, vol. 7, pp. 838-847, Mar. 2008.

[7] T. Ye, M. Walsh and B. OFlynn. “Experimental Analysis of Transmit-power for IEEE 802.15.4a UWB Ranging under Multipath Environment”. *ISSC*, NUI Maynooth, June 28-29, 2012.

[8] X. Li and K. Pahlavan. “Super-resolution TOA estimation with diversity for indoor geolocation”. *IEEE Trans. Inf. Theory*, vol. 3, no. 1, pp. 224-234, 2004.

[9] Y. Luo and C. L. Law. “A least-squares algorithm for multipath estimation using an UWB-IR link”. *in Proc. IEEE Wireless Commun.*, Netw. Conf., pp. 1-6, 2010.

[10] D. Chazan, M. Zakai, and J. Ziv. “Improved lower bounds on signal parameter estimation”. *IEEE Trans. Inf. Theory*, vol. IT-21, pp. 90-93, Jan. 1975.

[11] R. D. Yates. “A framework for uplink power control in cellular radiosystems”. *IEEE Journal on Selected Areas in Communications*, vol.13, pp.1341-1347,Sept.1995.

[12] M. J. Walsh and M. J. Hayes. “Robust Power Aware Mobile Agent Tracking using an 802.15.4 Wireless Sensor Network”. *IET Mechatronics*, 2008.

[13] M. Gabriella, D. Benedetto and G. Giancola, “Understanding Ultra Wide Band Radio Fundamentals”. *Prentice Hall*, 2004.

[14] www.decawave.com

 ISBN: 978-1-61208-296-7

MDS-based Algorithm for Nodes Localization in 3D Surface Sensor Networks

Biljana Risteska Stojkoska, Danco Davcev
Faculty of Computer Science and Engineering
University "Ss. Cyril and Methodius"
Skopje, Macedonia
biljana.stojkoska@finki.ukim.mk, danco.davcev@finki.ukim.mk

***Abstract*— As Wireless Sensor Network (WSN) has become a key technology for different types of smart environment, nodes localization in WSN has arisen as a very challenging problem in the research community. Most of the applications for WSNs necessitate a priory known nodes positions. In this paper, we propose an algorithm for three dimensional (3D) nodes localization in surface WSN based on multidimensional scaling (MDS) technique. Using extensive simulations, we investigated in details our approach regarding different network topologies, various network parameters and performance issues. The results from simulations show that our algorithm produces small localization error and outperforms MDS-MAP in terms of accuracy.**

Keywords-wireless sensor networks; multidimensional scaling; 3D surface localization; nodes positioning.

I. INTRODUCTION

A wireless sensor network (WSN) is a network of autonomous distributed sensor devices that obtain various measurements of different real-life occurrences [1][2]. After taking samples of physical or environmental conditions at different locations (light level, air temperature, humidity, etc.), each sensor sends data to its closest neighbor responsible for retransmitting the packets [3]. The final destination is the sink node responsible for storing data or for further processing. Although initially developed for military applications, today, WSNs are used in many industrial and civilian application areas, habitat monitoring, healthcare applications and traffic control [4].

Nodes localization is the basis for many applications of WSN, such as event detection and target tracking. A manual disposition is impossible not only for large scale WSNs, but also when a WSN is deployed on inaccessible terrain. The most straightforward solution to the localization problem is to apply Global Positioning System (GPS) receivers to each node [5]. But it is an expensive solution and inapplicable for indoor environments [6][7].

Finding out accurate positions of the WSN nodes without GPS support has been studied for many years. Many different techniques [6][7] have been proposed for solving this problem, but most of them consider only two-dimensional (2D) network. In this paper, we investigate multidimensional scaling technique [8] for nodes localization in three dimensional surface WSNs. We also propose a heuristic approach in distance matrix calculation that improves the accuracy compared with well known MDS-MAP [9]. Henceforth, we will refer to our approach as Improved Multidimensional Scaling Algorithm (IMDS).

The rest of this paper is organized as follows. The second section refers to the multidimensional scaling technique for nodes localization in 3D-WSN. The third section gives a detailed explanation of our IMDS algorithm. Section four presents the results provided from our simulations. Finally, we conclude this paper in section five.

II. THREE DIMENSIONAL MDS

Multidimensional scaling (MDS) is a set of analytical techniques that has been used for reducing the dimensionality of the data (objects), showing multidimensional data as points in 2D or 3D space [8]. MDS algorithm uses the distances between each pair of object as input and generates 2D-points or 3D-points as output. The input required by MDS should be presented as distance matrix, representing the distances between the objects that should be analyzed. The purpose of this method is to visualize dissimilarity data in order to better understand and comprehend it.

MDS can be easily translated into WSN domain if the sensor network and distances between neighboring nodes are represented as a graph with its edges respectively. In WSNs, MDS performs as centralized, range-based localization algorithm. Distance measurements between each pair of neighboring nodes will be collected at the sink node. There, all available information will be used in order to obtain the unknown distances between non-neighboring nodes.

There are a few well-known techniques for distance measurement between neighboring nodes [6][7][10], like Received Signal Strength Indicator (RSSI), Time of Arrival, Time Difference of Arrival (TDoA) and Angle of Arrival (AoA). RSSI [10] measurement of distances is often preferred as it does not require additional hardware. RSSI is based on the phenomenon that the intensity of emitted signal decreases as the distance from the signal source increases. If the function of the attenuation in dependence on a distance is known in advance, the distance between the emission source and the receiver can be easily estimated. The time needed for a message to travel from one node to another is used to provide range information in ToA and TDoA techniques, while AOA is defined as the angle between the propagation direction of the wave and some reference direction.

The main advantage of using MDS is its ability to reconstruct the relative map of the network even when there

 ISBN: 978-1-61208-296-7

are no anchor nodes (nodes with a priori known location). If given sufficient portion of anchor nodes, MDS performs very accurate position estimation enabling local map to be transformed into an absolute map [9][11].

There are different versions of MDS for nodes localization in a two dimensional WSN. The most popular is MDS-MAP, proposed by Yi Shang and Wheeler Ruml [9], where Dijkstra algorithm is used to calculate the unknown distances from the distance matrix. In [9] it is shown that MDS-MAP outperforms other techniques, especially when applied on density networks. Other approaches based on MDS-MAP exist [12], but most of them are complex and thus more computationally dependent. In [13], the authors introduce MDS-MAP(P), which is a decentralized version of the MDS-MAP. MDS-MAP(P) outperforms MDS-MAP on irregular network topologies, but requires intensive computational resources at each node. It computes local maps at each node in the network and then merges local maps into a global map. Using absolute positions of the anchors, this global map can be easily transformed into an absolute map.

Although a lot of research has been carried out regarding MDS-MAP for WSN localization, all of the algorithms proposed in the literature based on MDS-MAP consider only two dimensional networks. To the extent of our knowledge, this is the first research that extensively investigates three dimensional surface WSN localization based on MDS.

A. Multidimensional scaling (MDS) for 3D-WSN

MDS-MAP for 3D WSN consists of 3 steps:

- Step 1: Calculate shortest distances between every pair of nodes (using either Dijkstra's or Floyd's all pairs shortest path algorithm). This is the distance matrix that serves as input to the multidimensional scaling in step 2.
- Step 2: Apply classical multidimensional scaling to the distance matrix. The first 3 largest eigenvalues and eigenvectors give a relative map with relative location for each node.
- Step 3: Transform the relative map into absolute map using sufficient number of anchor nodes (at least 4).

B. Finding optimal rotation and translation between corresponding 3D nodes

Generating an absolute map (step 3) of the WSN requires at least four anchor nodes.

Let $P = \{p_1, p_{2,\ldots,} p_N\}$ and $Q = \{q_1, q_{2,\ldots,} q_N\}$ be two sets of corresponding nodes, where N is the number of anchor nodes in the WSN. We wish to find a transformation that optimally aligns the two sets in terms of least square errors, i.e., we seek a rotation matrix **R** and a translation vector **t** such that

$$(R,t) = \arg\min_{R,t} \sum_{i=1}^{N} \left\| (Rp_i + t) - q_i \right\|^2. \quad (1)$$

This transformation is also known as Euclidean or Rigid transformation, because it preserves the shape and the size.

There are many algorithms purposed in the literature that compute a rigid 3D transformation [14]. The most explored are based on Singular Value Decomposition (SVD), as it is known to be the most stable [15]. Finding the optimal rigid transformation with SVD can be broken down into the following steps:

- Compute the weighted centroids of both point sets

$$\overline{p} = \frac{1}{N}\sum_{i=1}^{N} p_i,\ \overline{q} = \frac{1}{N}\sum_{i=1}^{N} q_i, \quad (2)$$

- Compute the centered vectors

$$p_i' := p_i - \overline{p},\ q_i' := q_i - \overline{q},\ \text{i=1,…,N} \quad (3)$$

- Compute the 3x3 covariance matrix

$$H = P'Q'^T, \quad (4)$$

where P' and Q' are the 3xN matrices that have p_i' and q_i' as their columns, respectively.

- Compute the singular value decomposition

$$H = U\Sigma V^T, \quad (5)$$

The rotation we are looking for is then

$$R = VU^T, \quad (6)$$

- Compute the optimal translation as

$$t = \overline{q} - R\overline{p}. \quad (7)$$

C. Time complexity of MDS-MAP for 3D-WSN

In step 1, distance matrix construction using Dijkstra's or Floyd's algorithm requires $O(n^3)$, where n is the number of nodes in the network. In step 2, applying MDS to the distance matrix has complexity of $O(n^3)$ due to singular value decomposition. In step 3, the relative map is transformed through linear transformations. Computing the rigid transformation takes $O(N)$ time for computing P and Q, while computing SVD takes only $O(3^3)$ time (since the dimension of covariance matrix H is 3x3). Applying the transformation (rotation and translation) to the whole relative map takes O(n-N) time, where N is the number of anchors (N<<n).

III. IMPROVED MDS-BASED APPROACH FOR WSN POSITIONING

In this section, we will explain in details our improved multidimensional scaling algorithm (IMDS) for nodes localization in WSN.

MDS is very accurate technique for dimensionality reduction. If the correct distance matrix is given as input, MDS algorithm will reconstruct the map of the network without error. But, calculating distance matrix for networks

 ISBN: 978-1-61208-296-7

where only distances between neighboring nodes are known is not a trivial task. This problem in MDS-MAP is solved by applying Dijkstra's (or Floyd's) all pairs shortest path algorithm. Dijksta's algorithm is a graph search algorithm that solves the single-source shortest path problem. In WSN localization problem, the sensor network is represented as a graph with non-negative edge path costs, while the real, Eucledian distance between two non-neighboring nodes is replaced with the distance calculated using Dijksta algorithm. But the assumption that Dijkstra distance between two nodes correlates with their Euclidean distance is hardly true. This approximation produces an error, i.e., the positions obtained as MDS output usually differ from the correct positions. The difference between the real and the predicted positions is known as estimation error. The error is bigger when the nodes are in multi-hop communication range, which is a common case in obstructed environments. It is usually caused by the presence of obstacles or terrain irregularities that can obstruct the line-of-sight between nodes or cause signal reflections. Fig. 1 shows two examples when Dijksta algorithm will calculate much larger distance between non-neighboring nodes. Left side of the picture shows an example of two nodes A and B that are far from each other. The distance between A and B will be calculated as $AB=a+b+c+d$, which is much longer then the real Euclidian distance. This scenario is present when the network is deployed on vast regions where the radio range of the nodes is short compared with the length of the region. On the right side of Fig. 1, there is an example where two nodes can't communicate directly although they are very close to each other. The reason for this is the presence of obstacle that obstructs the line-of-sight. In this scenario, Dijksta algorithm is completely inapplicable as it calculates a few times longer distance.

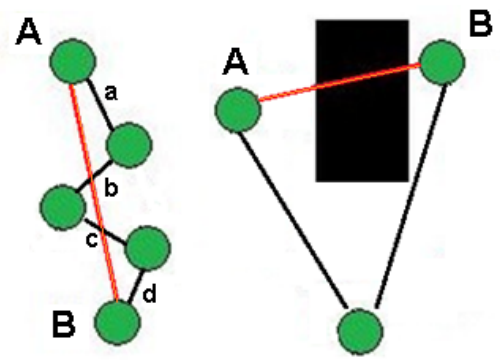


Figure 1. Distance approximation

As it can be seen from the two examples presented in Fig. 1, the distance calculated using Dijksta algorithm always increase the real distance. In order to reduce this distance, in this paper, we propose an alternative heuristic approach. By reducing the distance matrix error, we intend to reduce the overall estimation error.

A. *Distance matrix calculation*

Consider there are three nodes in a network: *A*, *B* and *C* (Fig. 2), with known distances between nodes *A* and *B* ($d_1=AB$), and between nodes *B* and *C* ($d_2=BC$). Since distance matrix requires the distances between every pair of nodes in the network, the distance between nodes *A* and *C* has to be obtained. We will refer to this distance as *a*.

If maximum radio range of the nodes in the network is *R*, then, we know for sure that node *C* can lay anywhere on the curve C_1C_2. If Dijkstra's algorithm is used for this purpose, it will calculate the distance *a* as $a=AB+BC$, which is the longest possible theoretical distance between nodes *A* and *C*. More precisely, *C* will lay exactly on C_2. On the other hand, if we calculate the shortest possible theoretical distance between nodes *A* and *C*, it will be very close to *R*. We can conclude that:

$$R < a \le d_1 + d_2 . \tag{8}$$

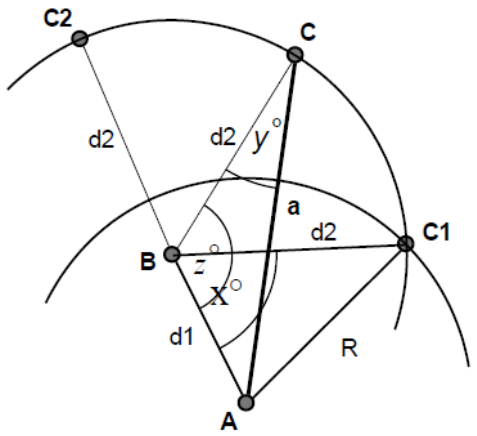


Figure 2. Distance approximation

To minimize the possible error, we purpose a heuristic solution that assumes that the node *C* lies exactly in the middle of the curve C_1C_2. Hence, the distance $a=AC$ can be calculated using cosine formula as:

$$a^2 = {d_1}^2 + {d_2}^2 - 2 \cdot d_1 \cdot d_2 \cdot \cos(\sphericalangle ABC). \tag{9}$$

In order to calculate the distance *a*, first, we need to find the angle using cosine formula:

$$\sphericalangle ABC = \sphericalangle ABC_1 + \sphericalangle C_1BC \tag{10}$$

The angle $\sphericalangle ABC_1$ can be calculated again with the cosine formula:

$$\sphericalangle ABC_1 = \arccos\left(\frac{{d_1}^2 + {d_2}^2 - R^2}{2 \cdot d_1 \cdot d_2}\right) \tag{11}$$

Since

$$\sphericalangle C_1BC = \sphericalangle CBC_2 , \tag{12}$$

$$\sphericalangle C_1BC = \frac{1}{2} \sphericalangle C_1BC_2 , \tag{13}$$

$$\sphericalangle C_1BC = \frac{1}{2} (\pi - \sphericalangle ABC_1), \tag{14}$$

 ISBN: 978-1-61208-296-7

$$\sphericalangle ABC = \sphericalangle ABC_1 + \frac{1}{2}(\pi - \sphericalangle ABC_1),$$

$$\sphericalangle ABC = \frac{\pi}{2} + \frac{1}{2}\sphericalangle ABC_1 \quad (15)$$

Finally,

$$a^2 = d_1^2 + d_2^2 - 2 \cdot d_1 \cdot d_2 \cdot \cos(\sphericalangle ABC) =$$

$$= d_1^2 + d_2^2 - 2 \cdot d_1 \cdot d_2 \cdot \cos(\frac{\pi}{2} + \frac{1}{2}\sphericalangle ABC_1) =$$

$$= d_1^2 + d_2^2 + 2 \cdot d_1 \cdot d_2 \cdot \sin(\frac{1}{2}\sphericalangle ABC_1), \quad (16)$$

where

$$ABC_1 = \arccos(\frac{d_1^2 + d_2^2 - R^2}{2 \cdot d_1 \cdot d_2}) \quad (17)$$

We note here that our algorithm preserves the time complexity of MDS-MAP algorithm.

IV. PERFORMANCE EVALUATION

The performance of the algorithms for WSN localization depends on different network parameters, such as the network topology, the number of anchors (i.e., the anchor-to-node ratio), the radio range, the density of nodes, etc. Hence, the location estimation error is going to be evaluated as a function of different parameters.

A. Network model

We assume a typical sensor network composed of hundreds (or thousands) of sensor nodes deployed uniformly across three dimensional monitored area (valley or mountain). Each sensor is equipped with an omni-directional antenna and has limited resources (CPU, battery, memory, etc.). Since radio signals are omni-directional, only nodes within certain radio range R can communicate with each other. If two nodes are within each others transmission range they are called neighbors. Further, we made following assumptions:

- Nodes are static and unaware of their location.
- There is a path between every pair of nodes.
- Nodes deployed in close proximity to each other exchange messages.
- Each node uses RSSI (or any other) method for distance estimation.
- RSSI provide accurate neighboring sensor distance estimation.

We simulated both techniques (MDS-MAP and IMDS) on different surface WSNs with Matlab.

We considered:

- Different network topologies:
  - 100 nodes randomly deployed on valley terrain (topology I)
  - 100 nodes randomly deployed on mountain terrain (topology II)
- 4, 6, 10 and 15 anchors for absolute map construction (for 3D rigid transformation SVD method was used)
- Different radio ranges (*R*) that lead to different average connectivity (average number of neighbors).
- Radio range error *er* (from *er*= 0%R to *er*=30% R with step 5% of R)

Thus 280 different networks were simulated (2 x 4 x 5 x 7) and each node location was discovered with both MDS-MAP and IMDS technique. The connectivity parameter and the estimation error for each scenario represent average over 30 trials for both algorithms. The average estimation error is normalized by the radio range R:

$$Error = \frac{\sum_{i=1}^{n} \text{distance}(pos_i^{(estimated)} - pos_i^{(true)})}{(n-N) \cdot R} \cdot 100\%, \quad (18)$$

where n is the number of nodes in the network, N is the number of anchor nodes, $pos_i^{(estimated)}$ is the estimated location and $pos_i^{(true)}$ is the true location of the i-th node.

B. Comparison of MDS-MAP and IMDS for 3D surface WSN

It is expected that MDS-based algorithms for WSN localization will not work well for such scenarios, basically because of multi-hop distance between each pair of nodes. Our improved heuristic approach presented in this paper is expected to achieve more acceptable accuracy.

Fig. 3 shows an example of two typical 3D surfaces. On the upper picture there is a surface, which represents a valley, while the lower surface represents a mountain. In our simulations, two scenarios are constructed to emulate a terrain with a valley and a terrain with a mountain. 100 nodes are deployed randomly with a uniform distribution over these two surfaces.

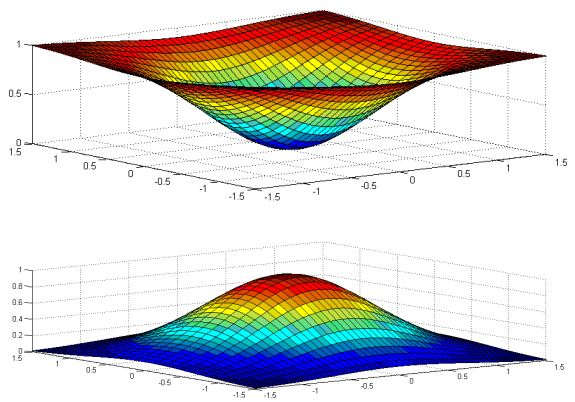


Figure 3. Typical 3D surface, valley (upper) and mountain (lower)

Fig. 4 and Fig. 5 compare the results of MDS-MAP and IMDS for topology I and topology II respectively.

 ISBN: 978-1-61208-296-7

In the case of topology I (Fig. 4), when *er* is small, both IMDS and MDS-MAP produce very similar estimation error. This error is much more affected by the number of anchors. As *er* increases, IMDS performs much better than MDS-MAP for all connectivity levels, regardless of the number of anchors.

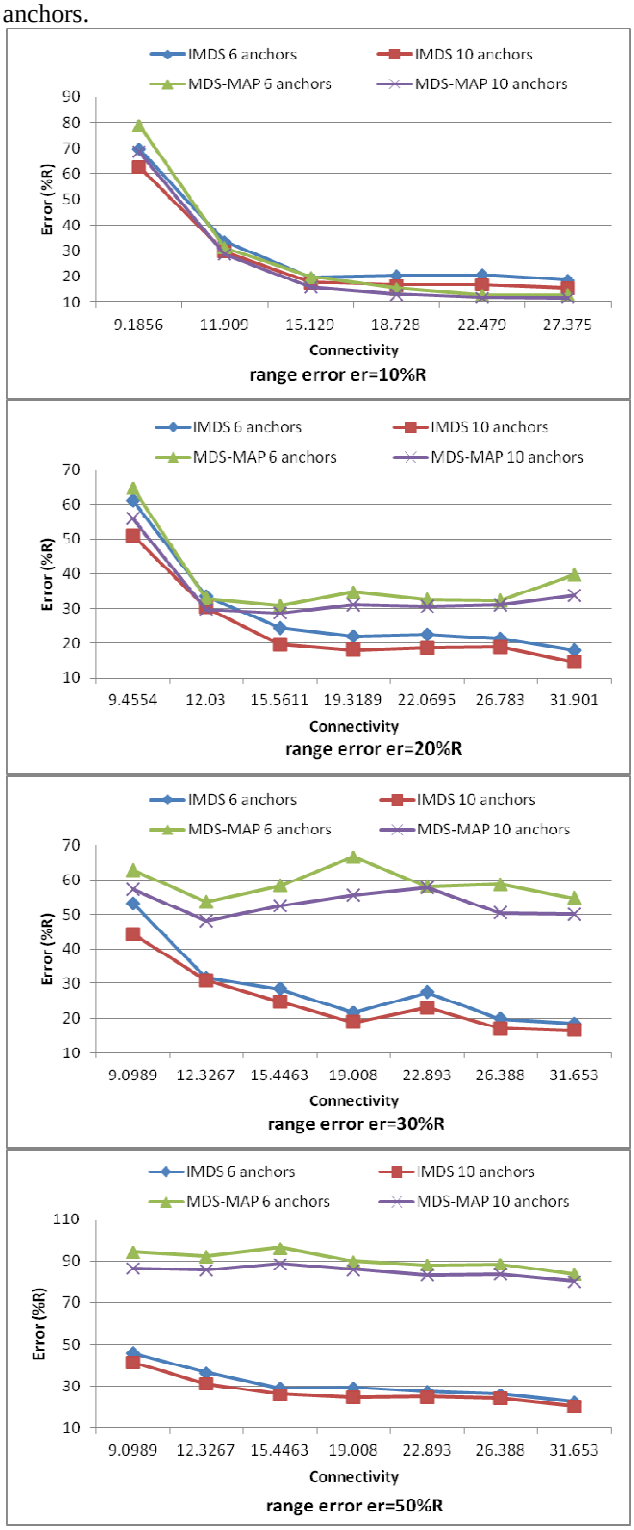


Figure 4. Comparison of MDS-MAP and IMDS for topology I

In case of topology II, for small *er* MDS-MAP has smaller estimation error than IMDS (Fig. 5). For large values of range error *er*, IMDS is better than MDS-MAP in terms of accuracy.

The characteristics of IMDS to produce smaller estimation error than MDS-MAP for large range error *er* is very important, as range measurement in the real applications is prone to error. When adopting distance measurement based on RSSI, the range error measurement is at least 10%R. The results presented in [10] show average range error measurement between 5%R and 30%R for longer radio range R. Similar research that investigates RSSI is conducted in [16] and [17], reporting average error around 20%R.

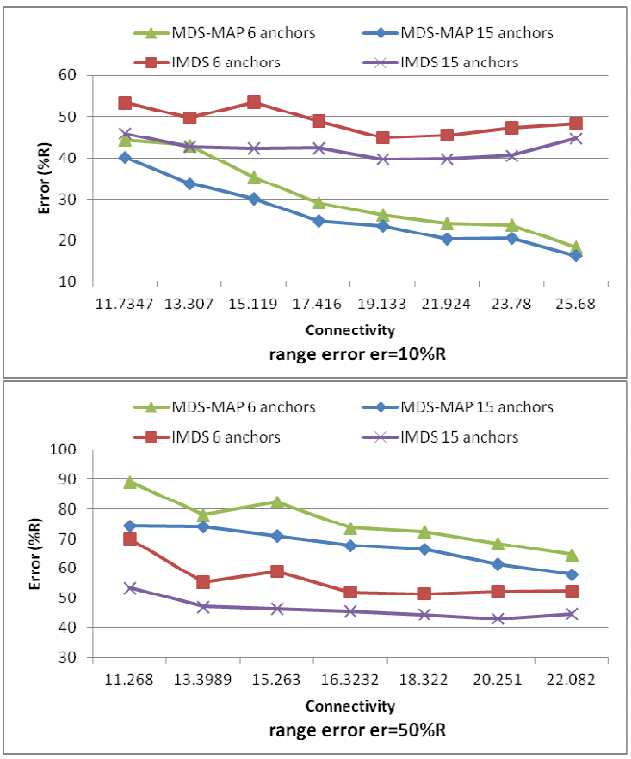


Figure 5. Comparison of MDS-MAP and IMDS for topology II

The average performance of IMDS as a function of connectivity for valley WSN is given on Fig. 6. IMDS is very stable and predictive. Estimation error decreases as connectivity increases. The radio range error *er* affects the estimation error in a way that larger *er* deteriorates the performance of IMDS.

As expected, the number of anchors affects the results, i.e., having more anchors slightly improves performance for all connectivity levels (Fig. 7).

If we compare the results for topology I and topology II, we can notice that both MDS-MAP and IMDS show better performance for topology I (valley terrain). The main reason for this is the characteristic of the terrain. Valley terrain is

 ISBN: 978-1-61208-296-7

very regular because all nodes that are within radio range R can communicate with each other. Mountain terrain should be considered as an irregular topology. The mountain presents an obstacle that obstruct the radio propagation between the nodes, which means that sometimes nodes that are very close to each other cannot communicate, i.e., cannot measure the distance between each other. For terrains with obstacles, nodes localization problem should be solved differently. IMDS algorithm should manage hierarchical network organization based on cluster formation. This cluster-based approach, which is already developed and implemented for 2D networks in [18], should be considered for 3D surface networks.

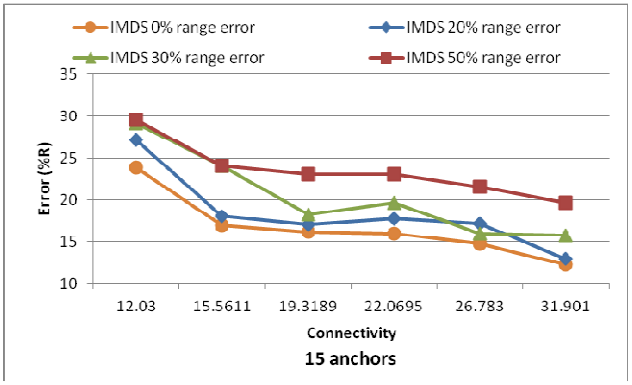


Figure 6. The effect of range error on the estimation error for topology I

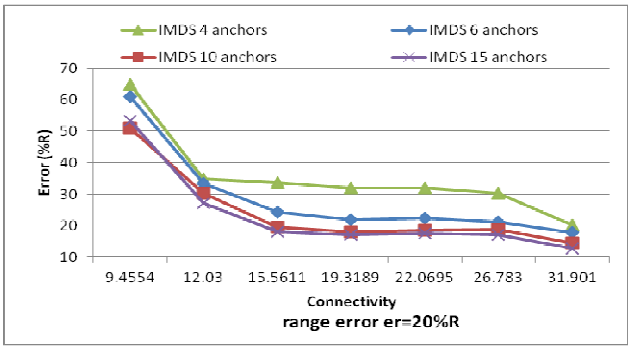


Figure 7. The effect of number of anchors on the estimation error for topology I

V. CONCLUSION AND FUTURE WORK

In this paper, we implemented improved MDS-based algorithm (IMDS) for nodes localization in 3D surface WSN. In IMDS, a novel technique for distance matrix refinement was introduced in order to reduce the estimation error. We investigated two surface network models (valley and mountain) and we showed that our approach outperforms MDS-MAP in terms of accuracy. IMDS performs much better than MDS-MAP especially when radio range error *er* is large.

For future work, we intend to investigate our algorithm on network where nodes are deployed on more complex 3D terrains. It is also a challenge to simulate radio propagation model in such complex 3D terrains, which is not a trivial task due to the presence of obstacles.

This way, we believe this work will contribute for future development of smart network technologies in different domains, especially for context- aware applications.

REFERENCES

[1] I. F. Akyildiz, S. Weilian, Y. Sankarasubramaniam, and E. Cayirci, A Survey on Sensor Network. IEEE Communications Magazine 40, 2002, pp. 102–114.

[2] F. L. Lewis, Wireless Sensor Networks, chapter 4 in D. J. Cook and S. K. Das, editors, Smart Environments: Technologies, Protocols, and Applications, John Wiley, New York, 2004.

[3] K. Akkaya and M. Younis, A survey of routing protocols in wireless sensor networks, Elsevier Ad Hoc Network Journal, May 2005, 3(3), pp. 325–349.

[4] C. F. Garcia-Hernandez, P. H. Ibarguengoytia-Gonzalez, and J. A. Perez-Diaz, Wireless Sensor Networks and Applications: A Survey, IJCSNS International Journal of Computer Science and Network Security, 7(3), (2007), pp. 264-273.

[5] B. W. Parkinson, 1994. GPS eyewitness: the early years. GPS World, 5(9), pp. 32-45.

[6] J. Wang, R. K. Ghosh, and S. K. Das, A survey on sensor localization, Journal of Control Theory and Applications, 8(1):2-11, 2010.

[7] A. Pal, Localization Algorithms in Wireless Sensor Networks: Current Approaches and Future Challenges. In Network Protocols and Algorithms, 2010, Vol. 2, No. 1, pp. 45-74.

[8] T. Cox and M. Cox, Multidimensional Scaling, Chapman & and Hall, London, 1994.

[9] Y. Shang, W. Ruml, Y. Zhang, and M. P. J. Fromherz, "Localization from mere connectivity," in Proceedings of ACM International symposium on mobile ad hoc networking and computing, June 2003, pp. 201-212.

[10] Z. Jianwu and Z. Lu, "Research on distance measurement based on RSSI of ZigBee," Computing, Communication, Control, and Management, 2009. CCCM 2009. ISECS International Colloquium on , vol.3, 8-9 Aug. 2009, pp. 210-212.

[11] X. Ji and H. Zha, "Sensor Positioning inWireless Ad-hoc Sensor Networks using Multidimensional Scaling," Proceedings of 23rd Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM 2004), March 2004, vol. 4, pp. 2662–2672.

[12] B. R. Stojkoska and V.Kirandziska "Improved MDS-based algorithm for nodes localization in wireless sensor networks," *IEEE EUROCON*, July 2013, pp. 608-613.

[13] Y. Shang and W. Ruml, "Improved mds-based localization," in Twenty-third AnnualJoint Conference of the IEEE Computer and Communications Societies (INFOCOM), 2004, pp. 2640– 2651.

[14] D. W. Eggert, A. Lorusso, and R. B. Fisher, "Estimating 3-D Rigid Body Transformations: A Comparison of Four Major Algorithms," Machine Vision and Applications (1997) Vol. 9, No. 5/6, pp. 272–290.

[15] A. Lorusso, D. Eggert, and R. Fisher, "A Comparison of Four Algorithms for Estimating 3-D Rigid Transformations", In: Proceedings of the 4th British Machine Vision Conference (BMVC 1995), Birmingham, England, (September 1995), pp. 237-246.

[16] A. Awad, T. Frunzke, and F. Dressler, "Adaptive distance estimation and localization in WSN using RSSI measures," in Proc. 10th Euromicro Conf. Digital Syst. Des. Architectures, Methods Tools (DSD) 2007, pp. 471–478.

[17] A. Faheem, R. Virrankoski, and M. Elmusrati, "Improving RSSI Based Distance Estimation for 802.15.4 Wireless Sensor Networks," in Conference Proceedings of the 2010 IEEE International

 ISBN: 978-1-61208-296-7

Conference on Wireless Technology and Systems (ICWIT 2010), Aug. 2010.

[18] B. Stojkoska, D. Davcev, and A. Kulakov, "Cluster-based MDS Algorithm for Nodes Localization in Wireless Sensor Networks with Irregular Topologies", Proceedings of The The Fifth ACM/IEEE International Conference on Soft Computing as Transdisciplinary Science and Technology (CSTST'08), Paris, France , October 26-30, 2008, pp. 384-389.

 ISBN: 978-1-61208-296-7

A Pre-Detection Query Tree Tag Anti-Collision Scheme in RFID Systems

Chiu-Kuo Liang, Yuan-Cheng Chien, and Chih-Hung Tsai
Dept. of Computer Science and Information Engineering, Chung Hua University
Hsinchu, Taiwan, R.O.C.
Email: {ckliang, e10002008, e09902009}@chu.edu.tw

***Abstract*—One of the research areas in RFID systems is a tag anti-collision protocol; how to reduce identification time with a given number of tags in the field of an RFID reader. There are two types of tag anti-collision protocols for RFID systems: tree based algorithms and slotted aloha based algorithms. Many anti-collision algorithms have been proposed in recent years, especially in tree based protocols. However, there still have challenges on enhancing the system throughput and stability due to the underlying technologies had faced different limitation in system performance when network density is high. Particularly, the tree based protocols had faced the long identification delay. Recently, a Hybrid Hyper Query Tree (H^2QT) protocol, which is a tree based approach, was proposed and aiming to speedup tag identification in large scale RFID systems. The main idea of H^2QT is to track the tag response and try to predict the distribution of tag IDs in order to reduce collisions. In this paper, we propose a pre-detection tree based algorithm, called the Pre-Detection Broadcasting Query Tree algorithm (PDBQT), to avoid those unnecessary queries. Our proposed PDBQT protocol can reduce not only the collisions but the idle cycles as well by using pre-detection mechanism. The simulation results show that our proposed technique provides superior performance in high density environments. It is shown that the PDBQT is effective in terms of increasing system throughput and minimizing identification delay.**

Keywords-Tag anti-collision; hybrid query tree; pre-detection query tree.

I. INTRODUCTION

Radio Frequency IDentification (RFID) is an emerging technology that guarantees to advance modern industrial practices in object identification and tracking, asset management, and inventory control [1]. Recently, several identification systems such as barcodes and smart cards are incorporated for automatic identification and data collection. However, these systems have several limits in read rate, visibility, and contact. RFID systems are a matter of grave concern because they provide fast and reliable communication without requiring physical sight or touching between readers and tags.

One of the areas of research is the speed with which a given number of tags in the field of RFID readers can be identified. For fast tag identification, anti-collision protocols, which reduce collisions and identify tags irrespective of occurring collisions, are required [1], [2], [3], [4], [5], [6], [7]. There are two types of collisions: reader collisions and tag collisions. Reader collisions indicate that when neighboring readers inquire a tag concurrently, so the tag cannot respond its ID to the inquiries of the readers. These collision problems can be easily solved by detecting collisions and communicating with other readers. Tag collisions occur when multi tags try to respond to a reader simultaneously and cause the reader to identify no tag. For low-cost passive RFID tags, there is nothing to do except response to the inquiry of the reader. Thus, tag anti-collision protocols are necessary for improving the cognitive faculty of RFID systems.

In general, the tag anti-collision techniques can be classified into two categories, aloha-based and tree-based protocols. Aloha-based approaches use time slot to reduce collision probability, such as Framed-Slotted aloha algorithm [1], [8], dynamic framed slotted aloha algorithm [5]. Tags randomly select a particular slot in the time frame, load and transmit its identification to the reader. Once the transmission is collided, tags will repeatedly send its id in next interval of time to make sure its id is successfully recognized. Aloha-based protocols can reduce the collision probability. However, they have the tag starvation problem that a particular tag may not be identified for a long time. For the consideration of performance, when number of RFID tag increased, the tag collision rate will be increased as well; this may result a low tag recognition rate.

The tree-based schemes use a data structure similar to a binary search algorithm, such as binary tree splitting protocol [3], query tree (QT) algorithm [9], and tree working algorithm [6], [10]. An RFID reader consecutively communicates with tags by sending prefix codes based on the query tree data structure. Only tags in the reader's interrogation zone and of which ID match the prefix respond. The reader can identify a tag if only one tag respond the inquiry. Otherwise the tags responses will be collided if multiple tags respond simultaneously.

Although tree based protocols deliver 100% guaranteed read rates, but they have relatively long identification delay. Recently, a hybrid query tree protocol (HQT) [11] was proposed and aiming to reduce transmission overhead by using 4-ary search tree mechanism and slotted backoff mechanism, in order to speed up tag identification and to increase the overall read rate and throughput in large-scale RFID systems. The main idea of the HQT technique is to reduce the number of collisions during the identification phase. In the 4-ary search tree mechanism, the prefix string of a collided query will be extended by 2-bits next time, unlike of 1-bit in the QT protocol. This way, collisions can be reduced substantially. Furthermore, the HQT protocol was aiming to reduce the idle cycles by using a slotted backoff

 ISBN: 978-1-61208-296-7

mechanism. When a tag responds to a reader, it sets its backoff timer using a part of its ID. If there is a collision (multiple tags respond), the reader can partially deduce how the IDs of tags are distributed and potentially reduce unnecessary queries.

Based on the HQT protocol, a H^2QT protocol [12] was proposed and aiming to reduce the idle cycles and improve the performance of tag identification. Although the H^2QT technique performs better than the HQT technique in reducing the number of idle cycles, it still has some idle cycles, which cannot be reduced during the tag identification process. In this paper, we proposed a PDBQT protocol to eliminate those unnecessary idle cycles. To evaluate the performance of our proposed technique, we have implemented our proposed PDBQT scheme along with previous proposed methods, HQT and H^2QT protocols. The experimental results show that the proposed technique presents significant improvement in most circumstance.

The remainder of this paper is organized as follows: Related work is discussed in Section II. In Section III, the tree based tag identification algorithm is introduced as preliminary of this study. In Section IV, our proposed algorithm, the PDBQT algorithm is presented. Performance comparisons and analysis of the proposed technique will be given in Section V. Finally, in Section VI, some concluding remarks are made.

II. RELATED WORK

Many research results for collision avoidance have been presented in literature. The existing tag identification approaches can be classified into two main categories, the Aloha-based [1], [4], [5], [8], [13] anti-collision scheme and the tree-based scheme [3], [6], [7], [9], [10]. RFID readers in the former scheme create a frame with a certain number of time slots, and then add the frame length into the inquiry message sent to the tags in its vicinity. Tags response the interrogation based on a random time slot. Because collisions may happen at the time slot when two or more tag response simultaneously, making those tags could not be recognized. Therefore, the readers have to send inquiries contiguously until all tags are identified. As a result, Aloha-based scheme might have long processing latency in identifying large-scale RFID systems [4]. In [1], Vogt et al. investigated how to recognize multiple RFID tags within the reader's interrogation ranges without knowing the number of tags in advance by using framed Aloha. A similar research is also presented in [14] by Zhen et al. In [13], Klair et al. also presented a detailed analytical methodology and an in-depth qualitative energy consumption analysis of pure and slotted Aloha anti-collision protocols. Another anti-collision algorithm called enhanced dynamic framed slotted aloha (EDFSA) is proposed in [5]. EDFSA estimates the number of unread tags first and adjusts the number of responding tags or the frame size to give the optimal system efficiency.

In tree-based scheme, such as ABS [3], Improved Bit-by-bit Binary-Tree (IBBT) [15] and IQT [16], RFID readers split the set of tags into two subsets and labeled them by binary numbers. The reader repeats such process until each subset has only one tag. Thus, the reader is able to identify all tags. The adaptive memoryless tag anti-collision protocol proposed by Myung et al. [2] is an extended technique based on the query tree protocol. Choi et al. also proposed the IBBT (Improved Bit-by-bit Binary-Tree) algorithm [15] in Ubiquitous ID system and evaluate the performance along three other old schemes. The IQT protocol [16] is a similar work approach by exploiting specific prefix patterns in the tags to make the entire identification process. Recently, Zhou et al. [17] consider the problem of slotted scheduled access of RFID tags in a multiple reader environment. They developed centralized algorithms in a slotted time model to read all the tags. With the fact of NP-hard [17], they also designed approximation algorithms for the single channel and heuristic algorithms for the multiple channel cases.

Although tree based schemes have advantage of implementation simplicity and better response time compare with the Aloha based ones, they still have challenges in decreasing the identification latency. In this paper, we present an enhanced tree based tag identification technique aims to coordinate simultaneous communications in large-scale RFID systems, to speedup minimize tag identification latency and to increase the overall read rate and throughput. Simulation results show that our proposed technique outperforms previous techniques.

III. TREE BASED TAG ANTI-COLLISION SCHEMES

In this section, we present three tree-based anti-collision techniques, namely the QT algorithm [4], the HQT algorithm [11], and the H^2QT algorithm [12], that are most related to our work.

A. Query Tree algorithm

The QT algorithm uses binary splitting strategy to identify tags. A reader transmits the *k*-length prefix. Then tags send from $(k + 1)th$ bit to the end bit of tag IDs if the first *k* bits of tag IDs are the same as the prefix. Also, if the received tag IDs collide, the extended prefix attached '0' or '1' to the prefix is retransmitted. Furthermore, if there is no collision, the reader identifies one of the tags.

B. Hybrid Query Tree algorithm

In the environment with high tags density, collision may happen very frequently while using the query tree algorithm, and due to that, a lot of query time will be wasted. By using the 4-ary search query tree mechanism, HQT can enable the prefix to increase two bits at a time from 1 bit. In this way, some collisions occurred in QT protocol can be reduced in HQT protocol. However, the drawback of the 4-ary search tree mechanism is the increasing number of idle cycles. To resolve this problem, HQT protocol introduces the slotted backoff mechanism. The slotted backoff mechanism is a technique that makes tags respond to the transmit prefix after waiting a certain time, instead of immediately respond. When a tag responds to a reader, it sets its backoff timer using a part of its ID. The backoff time of each tag is determined from the 2-bits, which follow the prefix of tag ID identical to the query prefix string. For example, tags do not defer their response if it is '00'. If it is '01', '10', or '11', tags will defer 1, 2, or 3 backoff time slots until they

 ISBN: 978-1-61208-296-7

respond to the reader, respectively. Fig. 1 shows the operation of the slotted backoff mechanism in HQT algorithm.

C. Hybrid Hyper Query Tree algorithm

The HQT algorithm, using slotted backoff mechanism, tries to reduce the number of idle cycles. By checking the starting and end point, HQT algorithm can reduce unnecessary expansion. However, only the starting and end idle slots can be checked by HQT algorithm, those idle cycles between busy slots cannot be reduced. To resolve the problem, the H^{2}QT algorithm uses a different slotted backoff mechanism. The backoff time of each tag is determined from the 3-bits, which follows the prefix of tag ID identical to the prefix. Unlike the mechanism used in HQT, the H^{2}QT counts the number of '1' in the following 3-bits and uses this number as the selected time slot for tags to respond. Fig. 2 shows the tag selecting its response slot based on tag ID.

Fig. 3 depicts an example of the query tree structure of identifying 5 tags with 6-bits ID length using H^{2}QT algorithm. The process of the identification is as follows: First of all, the reader sends request command with the empty-prefix to the tags. In this case, tags A, C and D will delay one time slot to respond since the first 3-bits of their tag IDs contain only one '1', as shown in Fig. 4. Similarly, tags B and E will delay two time slots to respond due to the number of '1' in the first 3-bits if their tag IDs is 2. In this case, since no tag responds immediately, it means that there is no tag whose first 3-bits of their tag IDs match '000'. Therefore, there is no need for reader to send the prefix string '000'. Similarly, since no tag responds after 3 time slot delay, the reader does not need to send the prefix string '111'. Therefore, the idle cycles can be eliminated.

Next, the reader receives tag IDs from tags A, C and D after one time slot delay. At this moment, the reader is aware that the pattern of the first 3-bits of tags A, C and D is 'X0X', in which 'X' represents a collision bit. Thus, the reader recognizes that the first 3-bits of tag A, C and D may be '001' or '100', which will be added into the queue for re-transmission. Similarly, the reader is aware that the bit pattern of the first 3-bits of tags B and E is 'XX1' after two time slots delay. Thus, the reader will put prefix strings '011' and '101' into the queue for re-transmission.

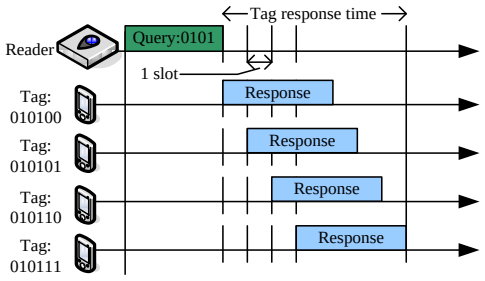


Figure 1. Operation of slotted backoff mechanism in HQT algorithm

Next, the reader then sends the request command with prefix string '001'. At this moment, only tag A responds after 1 time slot delay. In this case, tag A is identified by the reader. Table 1 summarizes the detail steps of communication between the reader and the tags with the example shown in Fig. 3.

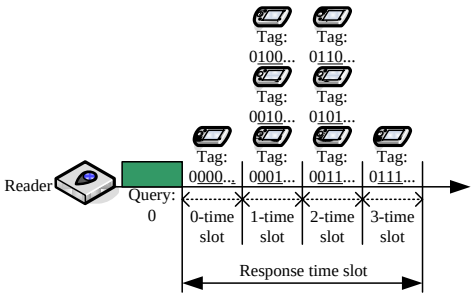


Figure 2. H^{2}QT response algorithm

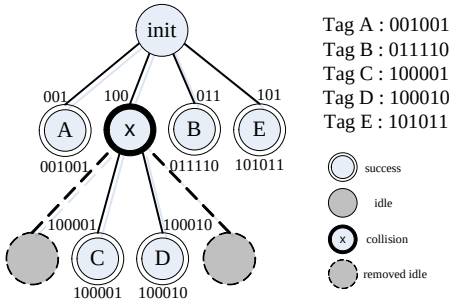


Figure 3. An example of H^{2}QT algorithm

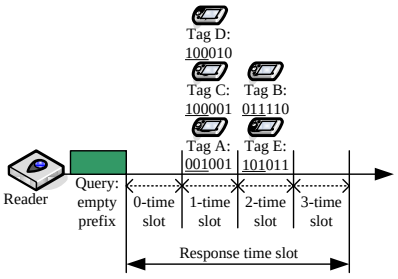


Figure 4. The response of tags in Fig. 3 after reader's empty-prefix request

TABLE I. COMMUNICATION STEPS OF FIG. 3.

Step	HQT		H^{2}QT	
	Broadcast	Status	Broadcast	Status
1	empty	Collision	empty	Collision
2	00	Identify Tag A	001	Identify Tag A
3	01	Identify Tag B	100	Collision
4	10	Collision	011	Identify Tag B
5	1000	Collision	101	Identify Tag E
6	1001	Idle	100001	Identify Tag C
7	1010	Identify Tag E	100010	Identify Tag D
8	100001	Identify Tag C		
9	100010	Identify Tag D		

 ISBN: 978-1-61208-296-7

IV. THE PROPOSED TECHNIQUE

Recall that, in H^2QT algorithm, the idle cycles can be reduced substantially. However, there still have some collision time slots. As a result, the reader has to spend more time slots to resolve the collisions. Due to that, it will take more time to complete the tag identification process. In this paper, we proposed a pre-detection scheme to eliminate the collision time slots and idle cycles.

A. Pre-Detection Broadcasting Query Tree Algorithm

We proposed a PDBQT algorithm, which uses pre-detection technique to realize the precise distribution of tag IDs. Once the distribution of tag IDs has been obtained, the reader broadcasts such message to tags and each tag is aware of the exact time slot to response. As a result, tags respond to the reader in different time slots and collisions can be avoided. Furthermore, since each tag realizes its corresponding time slot to respond, no empty time slot exists.

In our proposed PDBQT algorithm, after the reader send the request command to tags, the operations during the tag response period can be partitioned into three phases: the pre-detection phase, the broadcasting phase and the tag response phase, as shown in Fig. 5. The purposes of three phases design can be explained as follows: In pre-detection phase, the reader can realize the distribution of tag IDs by collecting the responses from tags. Then, in broadcasting phase, the reader will send the distribution information to tags so that each tag is aware of the time slot to send its ID to reader. Finally, in the tag response phase, the responses from tags are arranged into a sequence of time slots so that collisions and empty slots can be avoided.

In pre-detection phase, in order to collect the response information from tags, we allocate four short time slots for tags to respond, namely the ’00, ’01’, ‘10’, and ‘11’ time slots respectively. We also adapt the 4-ary search tree mechanism such that each tag whose tag ID matches with the prefix string sent from the reader will respond on the time slot according the following 2-bits of its tag ID. It should be noticed that, in this phase, each tag responds a 4-bits random number (RN) to reader instead of the whole tag ID. The reasons for tags of using 4-bits random numbers to respond are as follows: First, it can reduce the time for reader to realize the distribution of tag IDs, compared with the response of whole tag IDs. Second, the status of each time slot can be precisely identified with high probability. If no tag responds in a time slot, then the reader can correctly identify such time slot as an idle cycle, which can be eliminated during the tag response phase. If only one tag responds, the reader can also correctly identify such time slot as a successful cycle. Therefore, the reader will allocate a time slot to receive the response from that tag in the tag response phase. If more than one tag responds, since the tags respond 4-bits random numbers, a collision cycle can be identified by the reader by checking the received different random numbers. Although, there still has some chance for a reader to receive the same random number from different tags, however, the probability of successful collision detection is very high. Therefore, by using our pre-detection mechanism, the distribution of tag IDs can be correctly obtained with high accuracy.

Meanwhile, the reader monitors and records the response status from tags in each time slot during the pre-detection phase. The reader uses a ‘0’-bit to represent the time slot when no tag responds or more than one tag respond. On the other hand, the reader uses a ‘1’-bit to represent the time slot when only one tag responds. Therefore, after the pre-detection phase, the reader can use a 4-bits string to represent the status of four time slots in the pre-detection phase. Then, during the broadcasting phase, the reader broadcasts the 4-bits string to tags and by receiving the binary bit string, each tag can realize the exact time slot to respond by counting the number of ‘1’ in the received binary bit string from the start bit to its corresponding bit. Then, the tag can respond its tag ID to reader in the tag response time slot by finding the correct time slot to respond. For example, in Fig. 5, the tag which responds on the ‘11’ time slot in the pre-detection phase can realize that it can only send its tag ID on the third time slot in the tag response phase since it receives the binary bit string ‘1101’ sent from the reader and there are three ‘1’s from the beginning to its corresponding ‘1’.

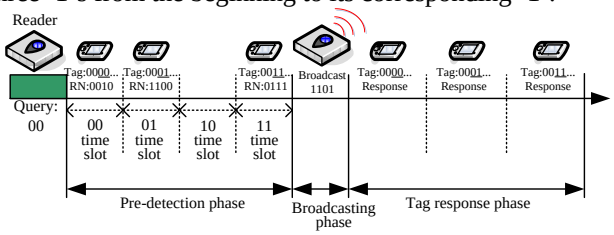


Figure 5. The tag response cycle of our proposed PDBQT algorithm

B. An Example

To facilitate the understanding of our proposed algorithm, an example is given as follows.

Fig. 6 depicts the example of the process of identifying 5 tags with 6-bits of tag IDs, ‘001001’, ’011110’, ‘100001’, ‘100010’ and ‘101011’, respectively, by using PDBQT protocol. The process of identification is as follows: First of all, the reader sends the request command with the empty-prefix to the tags. In this case, all tags respond to this request command and the time slot for a tag to respond is depending on the first 2-bits of its tag ID. In this example, tag A will respond in ‘00’ time slot, tag B will respond in ‘01’ time slot, tags C, D, and E will respond in ‘10’ time slot, and no tag responds in ‘11’ time slot, as shown in Fig. 6(a). Suppose that the random numbers for tags A, B, C, D, and E to respond are ‘1110’, ‘1100’, ‘0011’, ‘1010’, and ‘0111’, respectively. It can easily be seen that, since there is only one tag response for both ‘00’ and ‘01’ time slots, the reader will mark both time slots as ‘1’. Furthermore, since it has more than one tag responses in ‘10’ time slot, the reader will mark ‘10’ time slot as ‘0’. It should be noticed that as the reader recognize the collision time slot, the corresponding prefix bit string will be added into a queue for further requesting. In this example, the ‘10’ bit string will be added into the queue. After the pre-detection phase, the reader will mark all time slots as ‘1100’ and broadcast it to tags. After tags receive the message, tags A and B realize their own time slot to respond.

 ISBN: 978-1-61208-296-7

Therefore, tags A and B will be identified subsequently. In the meantime, tags C, D, and E recognize that the status of their time slot is '0', which means that they do not need to send their tag IDs to reader at that time slot, as shown in Fig. 6(a). After identifying tags A and B, the reader sends another request command from queue, which is the '10' bit string in this example as shown in Fig. 6(b). In this cycle, tag E is identified and bit string '1000' is added into queue. In the last round, as shown in Fig. 6(c), tags C and D can be identified.

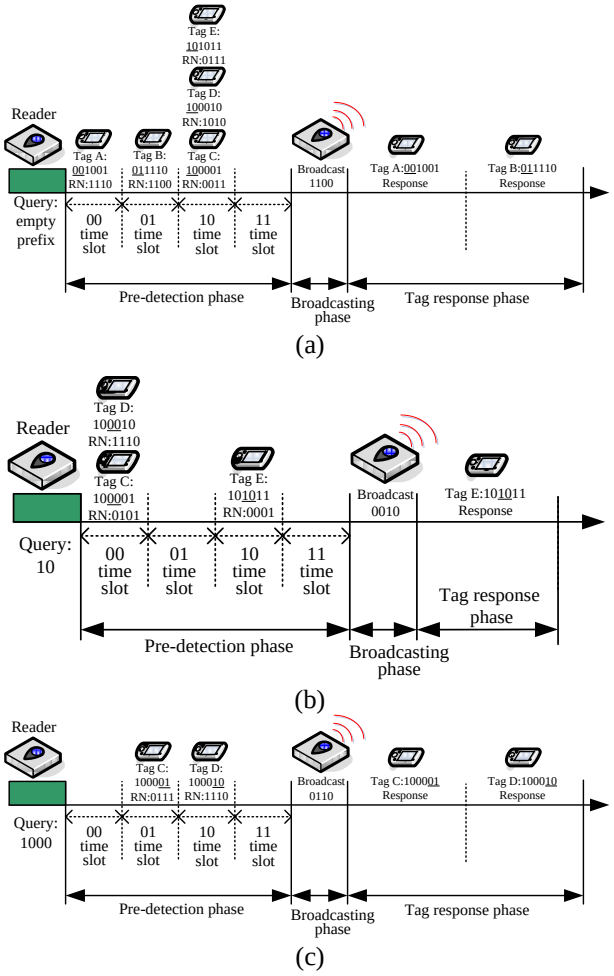


Figure 6. An example of identification process of our proposed PDBQT algorithm

C. Comparision of tag identification methods

To facilitate the understanding of the performance of our proposed algorithm, we compare the identification process between previous H^{2}QT and our proposed PDBQT algorithms by using the example in Fig. 6.

Table 2 shows required prefixes and steps for identifying all 5 tags by using different methods. In Table 2, the H^{2}QT scheme needs 7 steps to complete the identification process while in our proposed PDBQT scheme, only 3 steps are needed. Thus, our proposed PDBQT protocol reduces identification overhead efficiently and achieves better performance than H^{2}QT scheme.

TABLE II. COMMUNICATION STEPS OF FIG. 6.

Step	H^{2}QT		PDBQT	
	Broadcast	Status	Broadcast	Status
1	empty	Collision	empty	Identify tags A and B
2	001	Identify Tag A	10	Identify tag E
3	100	Collision	1000	Identify tags C and D
4	011	Identify Tag B		
5	101	Identify Tag E		
6	100001	Identify Tag C		
7	100010	Identify Tag D		

V. PERFORMANCE EVALUATION

To evaluate the performance of the proposed technique, we implemented the PDBQT scheme along with the HQT algorithm and the H^{2}QT algorithm. In the interrogation zone, we increase the number of tags from 500 to 4000. All tags are randomly generated in a uniform distribution manner. The lengths of the tag IDs used in each experiment are 96 bits. It should be noticed that some overhead are not taken into account in our simulation due to the communication latency and the propagation delay from the signal processing on the channel.

Fig. 7 shows the number of queries needed for reader to complete the tags identification. We can observe that, as the number of tags increases, each algorithm increases linearly due to the number of collision increases. However, our proposed PDBQT scheme requires less number of queries compared with other schemes.

Fig. 8 shows the number of idle cycles generated by each algorithm during the tag identification process. We can observe that, both H^{2}QT and our proposed PDBQT algorithm can eliminate all idle cycles regardless the number of tags increases.

Fig. 9 shows the number of collisions generated by each algorithm during the tag identification process. We can observe that our proposed PDBQT algorithm generates much fewer collisions than both HQT and H^{2}QT algorithms. Due to the pre-detection mechanism, most collisions can be detected in the pre-detection phase, there are only a few time slots wasted in the tag response phase.

Fig. 10 shows the total time required for each algorithm to complete the tag identification process. We can observe that our proposed PDBQT algorithm needs less time than both HQT and H^{2}QT algorithms to complete tag identification. Thus, the PDBQT algorithm outperforms the HQT and H^{2}QT algorithms.

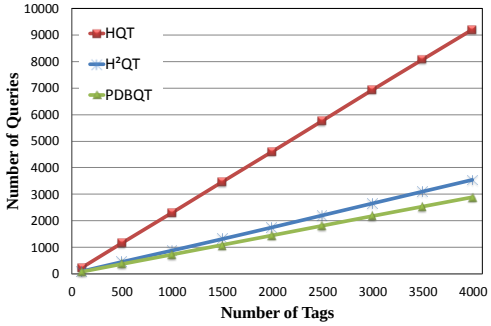


Figure 7. The number of queries required to complete tags identification.

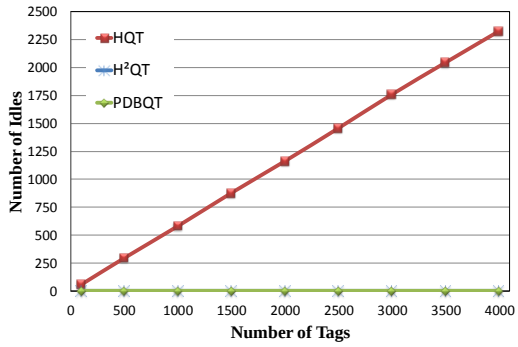


Figure 8. The number of idle cycles generated by each algorithm.

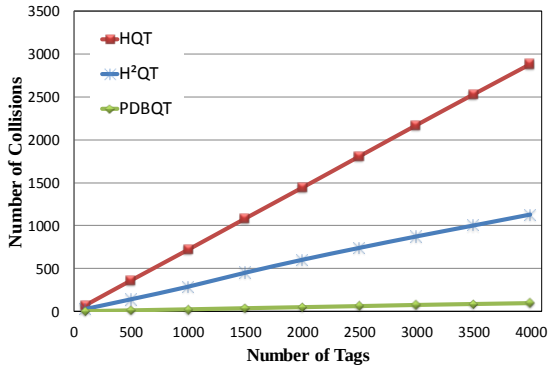


Figure 9. The number of collisions generated by each algorithm.

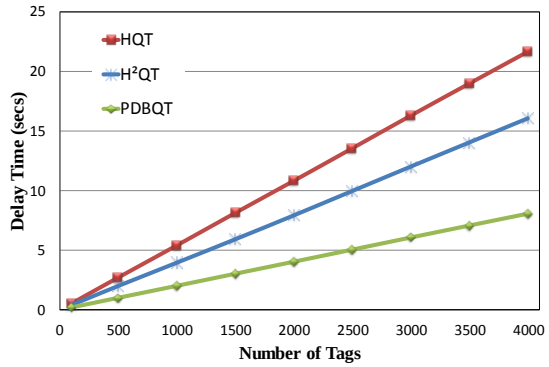


Figure 10. The time required to complete tag identification by each algorithm.

VI. CONCLUSIONS

With the emergence of wireless RFID technologies, identifying high density RFID tags is a crucial task in developing large scale RFID systems. Due to the nature of large scale RFID systems, many collisions may occur during the process of tag identification. In this paper, we proposed a nearly collision-free tag identification algorithm to reduce the iteration overhead efficiently. By using the random numbers for tags to respond in the pre-detection phase, many unnecessary collided inquiries can be reduced and the efficiency of tag identification can be significantly improved. To evaluate the performance of proposed techniques, we have implemented the PDBQT technique along with previous HQT and H^2QT algorithms. The experimental results show that the proposed technique provides considerable improvements on the latency of tag identification. It is also shown that the PDBQT is effective in terms of increasing system throughput and efficiency.

REFERENCES

[1] H. Vogt, "Efficient Object Identification with Passive RFID Tags," Proc. Inter. Conf. on Pervasive Computing, LNCS.2414, Springer-Verlag. Aug. 2002, pp. 98-113.

[2] J. Myung, and W. Lee, "An adaptive memoryless tag anticollision protocol for RFID networks", IEEE INFOCOM'05, Poster Session, Mar. 2005.

[3] J. Myung, W. Lee, and J. Srivastava, "Adaptive binary splitting for efficient RFID tag anti-collision," IEEE Communication Letter, vol. 10, no. 3, 2006, pp. 144–146.

[4] C. Law, K. Lee, and K. Y. Siu, "Efficient Memoryless Protocol for Tag Identification," in proceedings of the International Workshop on Discrete Algorithms and Methods for Mobile Computing and Communications, 2000.

[5] S. Lee, S. Joo, and C. Lee, "An enhanced dynamic framed slotted ALOHA algorithm for RFID tag identification," in Proc. MobiQuitous, July 2005, pp. 166-172.

[6] J. I. Capetanakis, "Tree algorithms for packet broadcast channels," IEEE Trans. Information Theory, vol. 25, Sep. 1979, pp. 505-515.

[7] F. Zhou, D. Jin, C. Huang, and M. Hao, "Optimize the Power Consumption of Passive Electronic Tags for Anti-collision Schemes," Proc. The 5th Inter. Conf. on ASIC, Vol. 2, Oct. 2003, pp.1213-1217.

[8] J. Park, M. Chung, and T.-J. Lee, "Identification of RFID Tags in Framed-Slotted ALOHA with Robust Estimation and Binary Selection," IEEE Communications Letters, vol. 11, no.5, 2007, pp. 452-454.

[9] "860MHz-930MHz Class I Radio Frequency Identification Tag Radio Frequency and Logical Communication Interface Specification Candidate Recommendation, Version 1.0.1," Auto-ID Center, Nov. 14, 2002.

[10] B. Feng, L. J. Tao, G. J. Bo, and D. Z. Hua. Id-binary tree stack anti-collision algorithm for RFID. IEEE Transactions on Computers, 2006, pp. 207–212.

[11] J. Ryu, H. Lee, Y. Seok, T. Kwon, and Y. Choi, "A Hybrid Query Tree Protocol for Tag Collision Arbitration in RFID systems," IEEE International Conference on Communications (ICC-07), 2007, pp. 24–28.

[12] T. H. Kim and S. J. Lee, "A Hybrid Hyper Tag Anti-Collision Algorithm in RFID System," ICACT 2009, pp. 1276-1281.

[13] D. K. Klair, K. W. Chin, and R. Raad, "An investigation into the energy efficiency of pure and slotted aloha based RFID anticollision protocols," in proceedings of the IEEE WoWMoM'07, June 18–21, Finland, 2007.

[14] B. Zhen, M. Kobayashi, and M. Shimizui, "Framed aloha for multiple RFID objects Identification," IEICE Trans. on Comm, E88-B(3), 2005, pp. 991–999.

[15] H. S. Choi, J. R. Cha, and J. H. Kim, "Improved Bit-by-bit Binary Tree Algorithm in Ubiquitous ID System," in proceedings of the IEEE PCM2004, Tokyo, Japan, Nov. 29-Dec. 3, pp. 696–703, 2004.

[16] A. Sahoo, S. Iyer, and N. Bhandari, "Improving RFID System to Read Tags Efficiently," IIT Bombay, June: KRSIT Technical Report, 2006.

[17] Z. Zhou, H. Gupta, S. R. Das, and X. Zhu, "Slotted Scheduled Tag Access in Multi-Reader RFID Systems", in proceedings of the IEEE International Conference on Networks Protocols (ICNP), pp. 61–70, 2007.

 ISBN: 978-1-61208-296-7

Passive SAW Based RFID Systems Finding their Way to Harsh Environment Applications

Alfred Binder, Gudrun Bruckner, Jochen Bardong
Carinthian Tech Research AG
Villach, Austria
alfred.binder@ctr.at

***Abstract*—While CMOS (Complementary Metal Oxide Semiconductor) transponders clearly take the lead in the high volume markets, SAW (Surface Acoustic Wave) transponders continuously develop in niche applications for special industrial needs. Based on a completely different physical principle, SAW transponders draw their potential for harsh environment applications from their inherent temperature stability, radiation hardness and their low energy consumption. Obviously, SAW transponders for harsh environment are no low cost solutions and are therefore used for tracking of high value industrial goods and additional sensor telemetry like temperature sensing. The present work explains aspects of SAW transponder design to meet harsh environment specifications for temperatures up to 400°C. This includes the improvement of interdigital transducer stability through an Al/Ti sandwich structure, packaging aspects and robust metal antenna designs. An overview of application examples is given for heavy industries like steel production, oil and gas exploitation and automotive varnishing lines. Finally, a research outlook for high temperature RFID (Radio Frequency Identification) solutions will be given targeting applications at 600°C and higher.**

Keywords - Passive SAW transponders, SAW RFID system, harsh environment, wireless sensor

I. INTRODUCTION

SAW based RFID systems exist for a while now. One brand is the "SOFIS" from Siemens. It is a system mainly used for railway applications with references back to 1995 [1]. Among other applications it is for instance installed in the wheelset diagnostic system in the Eurotunnel, France/UK. The "OIS-W" system from Baumer Ident (now Identec Solutions) is another SAW RFID system among the first systems on the market. It specializes on car varnishing lines and was one of the first transponders with a specified cycling temperature up to 200°C [2]. One of the first drawbacks of SAW RFID tags was the limited code size. This issue was addressed by RF SAW, Inc. with the proposal of a global SAW ID Tag comprising a 128-bit tag platform, realizing a 64 bit EPC tag including anti-collision, multi-level error detection and other features [3]. While SAW RFID systems outperformed CMOS RFID systems until the early 2000's in terms of range, this is not the case anymore through the development of UHF (Ultra High Frequency) systems, which have a range up to several meters necessary in logistic applications. CMOS RFID transponders have the major advantage of being programmed during operation. Further they have the possibility to run communication protocols and encryption algorithms. Last but not least CMOS RFID transponders are very price competitive due to the economics of scale. Although CMOS transponders are established in many industrial logistic applications there is a number of interesting harsh environment niches for SAW RFID transponders and it is the intention of this work to show the technical progress of these systems and their key applications done at Carinthian Tech Research AG. Section II presents the basic principle of SAW transponders. In Section III, typical specifications for harsh environment are discussed and necessary SAW transponder adaptations are explained to meet these specifications. Harsh environment applications that were supplied in the past few years are presented in Section IV. Section V concludes and gives an outlook on future research targets in the field.

II. FUNCTION

A RF interrogation signal is transmitted by a transceiver and picked up by the antenna of the passive SAW transponder. The interdigital transducer (IDT) converts the received signal into a surface acoustic wave by the converse piezoelectric effect. The SAW propagates towards reflectors distributed in a characteristic barcode-like pattern and is partially reflected at each reflector. The acoustic wave packets returning to the IDT are re-converted into electrical signals by the IDT and re-transmitted to the transceiver by the transponder antenna. This response contains information about the number and location of reflectors as well as the propagation and reflection properties of the SAW. Common transceiver concepts to read out SAW RFID transponder are FMCW (frequency modulated continuous wave) or FSCW (frequency step continuous wave) radar designs in the open ISM band at 2.4 GHz. The bandwidth of 80 MHz in this band allows for a reasonable resolution of the transponder response. The acquired frequency domain data from the CW radar are converted in a time domain signal using inverse fast Fourier transformation (IFFT), where the pulses caused by each reflector are further evaluated according to the code design [4]. There are various ways to implement a code design in a SAW RFID transponder. A pulse position code as shown in Fig. 1 was developed [5]. Each reflector is assigned to a time slot in a digit block. Each of the n blocks can contain m different reflector positions resulting in m^n different codes.

 ISBN: 978-1-61208-296-7

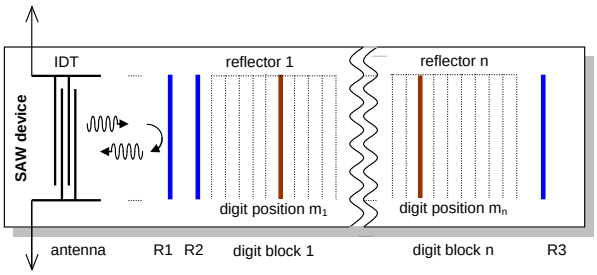


Figure 1. Pulse position code of SAW RFID transponder.

The reflectors R1, R2 and R3 are fixed and are used for optional temperature measurement. The SAW RFID transponder is processed on YZ-cut $LiNbO_3$ (LN YZ) due to its high electro-acoustic coupling coefficient and large TCD (temperature coefficient of delay) of -95 ppm/°C for temperature telemetry.

III. HARSH ENVIRONMENT SPECIFICATIONS

In industry "harsh environment" is a general term to describe working conditions, which are more stringent than what is considered as "normal". For this reason a detailed specification is needed for the exact working conditions. Generally, elevated temperatures and/or pressure in combination with mechanical abrasion, heavy shock loads or vibrations are to be considered. Some applications use high radiation doses or cryogenic temperatures. An exemplary overview of some typical harsh environment specifications is given below in Table I.

TABLE I. OVERVIEW OF HARSH ENVIRONMENT SPECIFICATIONS

Application / Industry	*Specification*
Tracking/Tracing of investment goods /steel industry	Temperature: 200°C – 300°C Long read ranges of several meters Mechanical collision, steel/slag splashes
Identification of slide-gate plates / steel industry	Temperature: 200°C – 380°C Shock loads and mechanical collision
Drill pipes / Oil and Gas	Temperature: 250 °C Pressure: 1300 bar Range: 0.5 m
Varnish line / Automotive	Temperature: 220 °C cyclic Range: several meters
Annealing of light alloy rim / Automotive	Temperature: 150°C - 540°C(max) Range: < 1 m
Autoclaving	Temperature: 200°C Pressure: 12 bar Saturated steam
Sterilisation (Gamma)	Food sterilization: 10 kGy Medical sterilization: 50 kGy
Cell banks / Life Science	Temperature: -196°C Range: 10 cm – 20 cm

In addition to harsh environment specifications of the transponder system specifications have often to be taken into account. Due to the hot environment reader antennas are exposed to temperatures exceeding 100°C. Often long RF cable lengths have to be installed as there is no space for reader electronics near to the reader antennas. Another practical problem is the heavy RF pollution of the ISM band in industrial environments. The SAW transceiver being limited to 10mW EIRP according to EN300440 contributes very little to this RF pollution. A typical industrial environment also contains a lot of metallic surfaces causing multipath propagation of the RF signal and even unwanted resonances. In some cases precaution has to be done using shielding or attenuation means.

IV. RUGGEDIZED SAW TRANSPONDER DESIGN

Generally, SAW devices are mass produced for the telecom market, but consequent cost optimizations lead to devices which are not suitable for harsh environment. Necessary adaptations are concerned with IDT metallization, packaging and antenna design, which are described below.

A. IDT Metallisation

Annealing experiments revealed that a single Al/Ti stack with a thickness of 50 nm degraded within 450 hours at 300°C [6]. The observed agglomeration is known to start with the Tammann temperature, which is defined as half the melting point of the bulk material (expressed in Kelvin). The thermal stability of the IDT metallization was significantly enhanced by putting two Al/Ti stacks (sandwich) on top of each other (Fig. 2). The Al/Ti sandwich has a total thickness of 70 nm and results in a tenfold better thermal stability.

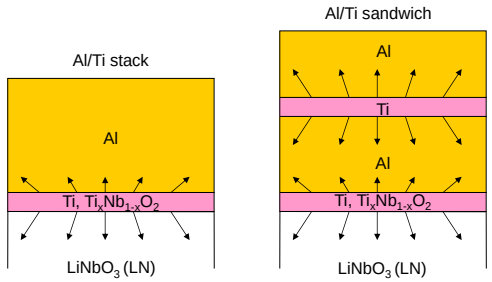


Figure 2. Scheme of the Al/Ti and Al/Ti sandwich metallisation.

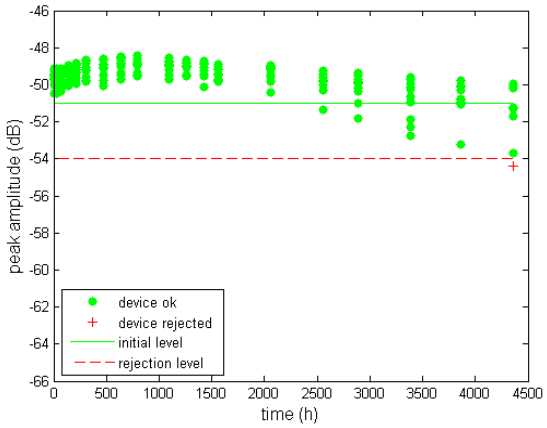


Figure 3. Amplitude of the minimum signal level after long-time annealing at 300°C.

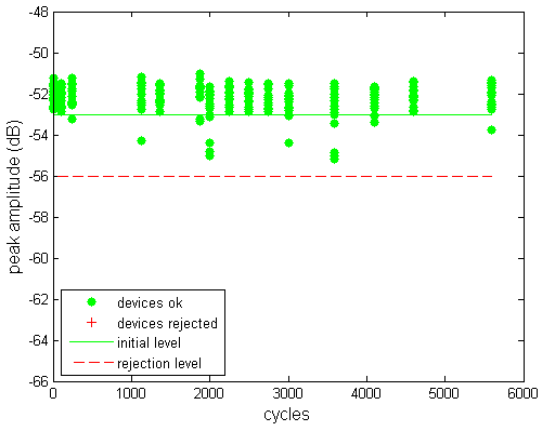


Figure 4. Amplitude of the minimum signal level after temperature cycling from 30°C to 230°C.

In the long-time annealing in Fig. 3, the devices show a run-in effect during the first 1000 hours, where the signal amplitude actually increases. After 4350 hours (more than 6 months) the first of ten devices dropped below the rejection limit of 3 dB. The effect of temperature cycling was tested by sequential cooling of 15 devices every 15 minutes from 30°C to 230°C. The development of signal amplitudes over cycles was measured. As can be seen from Fig. 4 even after 5600 cycles no device fell below the rejection limit. The deviations in the peak amplitude are presumably artifacts of the measurement (e.g., variations of the electrical contact resistance due to oxidation of contact pins).

B. Packaging

The bare SAW die is very sensitive to surface pollution. Thus it has to be hermetically sealed and protected with inert gas (N_2). Metal housings have a very good thermal stability, which is limited by the glass feed-through at approximately 450°C. The SAW is die-bonded to the metal housing using a silver-filled polyimide adhesive. Silicon based adhesives can only be used up to +250°C maximum. Ceramic adhesives were found to be inappropriate as they are too brittle to assimilate the non-uniform expansion of LN YZ substrate (13.4 ppm/°C along the a-axis and 4.1 ppm/°C along the c-axis). Wire bonding was performed using a 25 µm Al bond wire (wedge-wedge). The lid is welded to the socket using projection welding. Metallic housings were manufactured for different application purposes. A TO39 housing as shown in Fig. 5 can serve as a simple and cost effective housing for applications up to 400°C without further specifications on space requirements or mechanical pressure. On the other hand a TO39 housing can be used for cryogenic applications when the housing is enclosed in vacuum instead of nitrogen. A small custom housing with dimensions of 1.9 x 3.1 x 10.0 mm^3 was developed for applications that require a good thermal contact between tagged object and SAW transponder (Fig. 6). For high pressure applications like oil and gas exploitation a custom Inconel housing was designed (Fig. 7).

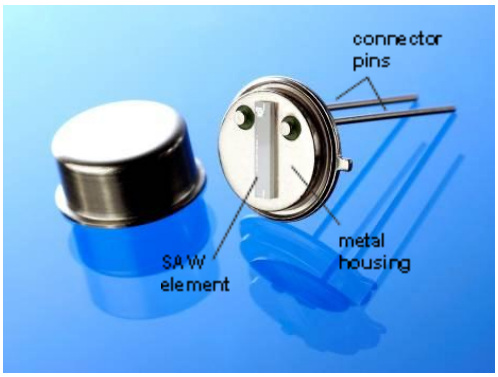


Figure 5. Transponder housing TO39, diameter 10 mm x 4 mm.

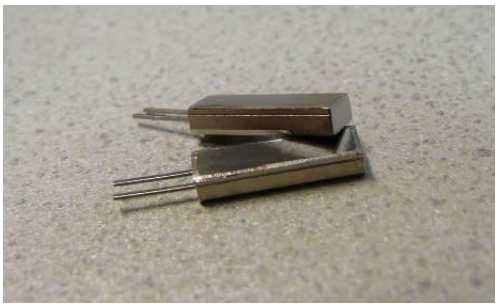


Figure 6. Custom Kovar housing 1.9 x 3.1 x 10.0 mm^3.

Figure 7. Custom Inconel housing for high pressure applications, diameter 9.1 mm x 5.1 mm.

The basic design was taken from TO39 standard, but all wall thicknesses were increased to withstand pressure up to 1300 bar. The diameter of the feed-through was changed from 1.5 mm to 1.3 mm to reduce the pressure on the glass.

C. Antenna design and transponder assembly

The antenna has to operate in the same working conditions as the transponder. For operation temperatures exceeding +250°C full metal antennas were designed. The electrical interconnection between transponder and antenna is laser welded, providing better temperature stability compared to soldering. Another aspect of the antenna design is to protect the transponder from mechanical shock loads and vibrations. This has been achieved by the development of a ruggedized slot antenna (Fig. 8).

 ISBN: 978-1-61208-296-7

Figure 8. Ruggedized, stainless steel slot antenna 100 mm x 40 mm with a thickness of 5 mm.

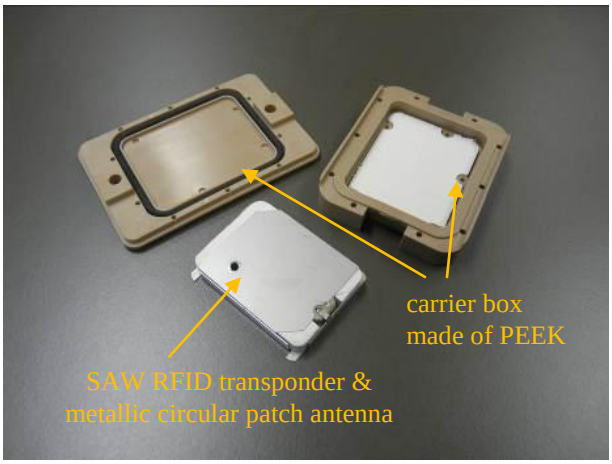


Figure 9. Metal circular patch antenna and PEEK housing, assembled: 95 mm x 64 mm x 17 mm.

The antenna is made of 5 mm thick stainless steel and is typically mounted on metal surfaces, which act as a reflector to increase the signal gain. As transponder housing a standard TO39 is used. The TO39 is well protected inside a cavity of the antenna. Whereas the slot antenna is linear polarized, a circular polarized metallic patch antenna has been developed. Fig. 9 shows the patch antenna, which is placed in between a carrier box made of PEEK (Polyether ether ketone). The metal patch antenna replaces a former design based on a PCB patch antenna and solder connections. The solution with metal antenna and welded electrical connection proved to be much more stable in temperature cycling conditions up to 220°C.

V. APPLICATIONS

The applications presented below give an overview of implemented RFID installations and pilots carried out in the past five years. The main asset of passive SAW RFID transponders in all those cases were the operability at temperatures >200°C, meaning that tags do not only withstand, but can also be read out at higher temperatures.

A. RFID in steel plant

RFID applications in the steel plant are motivated by the logistics optimization of moving investment goods or consumables. One of the main topics is the retraceability of the steel grade or the slag as a by-product for further process treatment [7]. Sometimes the transmitted temperature of the SAW tag is interesting add-on information.

1) Identification of a steel ladle

SAW RFID tags have been installed on the steel coating of steel ladles. The tags are positioned with a distance to the steel coating not to be in direct contact with the outer temperature of +300°C to +400°C. The tags are protected against slag splashes with refractory material on the outside (Fig. 10). The ladle is tracked along 10 different process positions in the steel plant. As the ladle moves along the plant the range between installed reader antennas and tag varies between 2 m and 5 m. At 5 m range stability problems of the SAW transponder read-out were reported. In the ongoing test, SAW RFID tags were long-term stable for several months. Through automated ladle tracking, process logistics is optimized, security is increased and a better retraceability achieved.

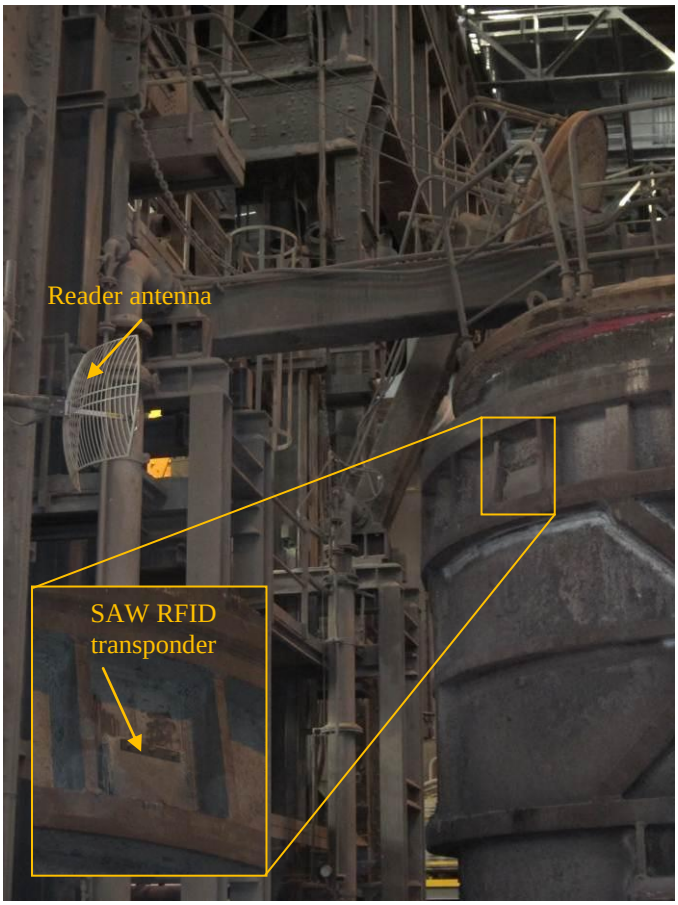


Figure 10. SAW RFID installation on a steel ladle (Georgsmarienhütte, Germany).

2) Identification of slide gate plates

A slide gate is steel equipment used to control the steel flow during casting. The core element is a sliding valve made of two plates (Fig. 11). These slide gate plates are refractory wear elements. Being crucial parts exposed to high temperatures, they must be replaced before a critical degree of wear has been reached. Failure (e.g., breakthrough of liquid metal) can cause great damage. An RFID and temperature monitoring system gives the manufacturer better control over the usage and the logistics are documented. Although at the center of the slide gate plates the temperature equals 1540°C, the melting temperature of steel, the temperature rapidly decreases towards the edges. A SAW transponder was installed on the outer rim (Fig. 12) of the plate. It faces a cavity in the slide gate mechanics, where a ruggedized reader antenna was mounted. The temperatures during casting were moderate around 250°C as cooling was applied. The maximum temperature was reached with 320°C on the maintenance station, when cooling is switched off. During a pilot test of several weeks about 100 slide gate plates were tested. All SAW transponders remained functional over all casts. In average a slight signal degradation of 2 dB to 5 dB was observed [8].

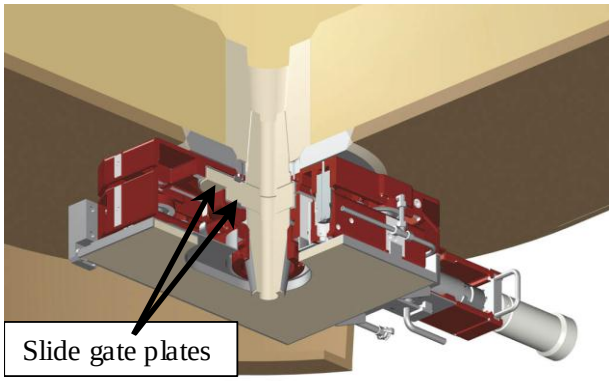


Figure 11. Cut of a slide gate mechanics to control the steel flow during casting (RHI AG, Austria).

Figure 12. SAW RFID installation on slide gate plate (RHI AG, Austria).

B. Asset tracking in the oil and gas industry

CMOS RFID transponder for down-hole equipment which are based on glass encapsulated low frequency (LF) transponder withstand the operating conditions, but have small read ranges. Applications like the automatic down-hole asset monitoring [9] call for significantly longer read ranges > 0.5 m. For a deep well specification temperatures up to 250°C and pressures up to 1300 bar have to be considered. A custom Inconel housing (Fig. 7) was developed to meet these requirements. The Inconel housing is connected to a spiral antenna and encapsulated in PEEK. The PEEK radome has a sacrificial layer to allow for wear on the tool in down-hole operation. The SAW RFID transponder is mounted in a bore hole with a depth of 16 mm via a press fit (Fig. 13). Although first down-hole tests with a few tags are promising a test with a representative number of tags is still due. The oil and gas industry demands a RFID solution for drill pipes to be applicable in static and dynamic cases for life-cycle management and asset predictability. The latter is more difficult as the pipes have to be read in and out dynamically at the rig. A rig floor reader has been designed which consists of a circular arrangement of reader antennas around the pipe. As the drill pipes are inserted the reader identifies them at arbitrary angular position. By tracking the pipe usage time and known parameters as well depth, well type and drilling method a fatigue prediction for each pipe can be derived. The benefit of this database is to preclude drilling down-time and fishing costs due to pipe breakage. Also a load-dependent maintenance of drill pipes can be implemented.

C. Automotive varnishing line

In automotive paint shops the varnish has to be cured at elevated temperatures >200°C. As different car bodies run through the paint shop each body has to be identified to start the correct painting process. Starting in the bodyshell work the SAW RFID transponder is mounted. At the assembly

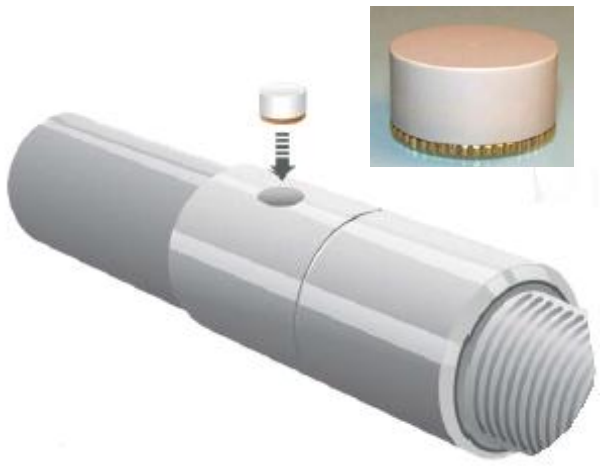


Figure 13. SAW RFID inserted into drill pipe collar/tool joint (HM Energy, LLC, USA).

 ISBN: 978-1-61208-296-7

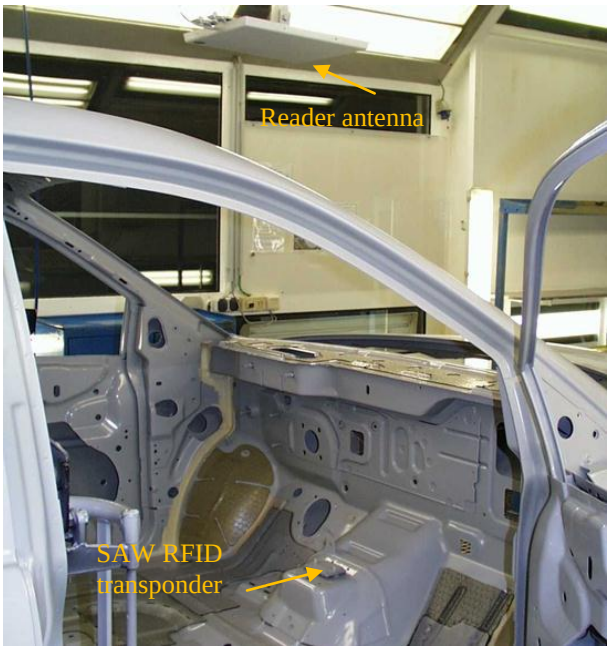


Figure 14. SAW RFID in an automotive varnishing line (Identec Solutions, Austria).

stands in the bodyshell work, before and after the paint shop and at some working areas in the end assembly the transponder is read out. In this way the car body is identified, processes are triggered and the car is consistently tracked within the networked assembly stands. The SAW RFID transponder according to Fig. 9 is attached to each car body. The transponder does not remain on the car body, but is recycled in the process. Fig. 14 shows the mounting of the transponder and in 1.5 m distance the read-out antenna. The main advantage of SAW RFID in this application is that the transponder offers a long lifetime and further the code can be read at elevated temperatures. One of the drawbacks of using SAW is the fixed code, which cannot be changed in the field. This has been solved using a code conversion table on the host side.

VI. RESEARCH OUTLOOK FOR A 600°C SAW RFID TRANSPONDER

State of the art SAW RFID transponder surpass the operation temperature of CMOS RFID transponder, but the SAW technology offers more potential to increase the current operation limit of +350°C to +400°C by far. Key elements of a high temperature SAW transponder that works beyond 600°C are the availability of a suitable substrate and IDT metallization. Aubert, et.al., have experimentally studied a Ir/Ti/AlN/Sapphire structure that exhibits a good stability of up to 1050°C for several hours [10]. The measurements were carried out on 392 MHz devices. The main degradation which occurred was due to agglomeration of the Iridium electrodes. Based on this very promising result, future research has to focus on the design and verification of 2.4 GHz devices. The ISM band at 2.4 GHz is a good choice due to the legal availability of 80 MHz bandwidth needed for pulse position coding. A second research focus has to deal with a suitable high temperature package. Metallic packages fail at 450°C because of the glass feed-through. Either a brazed ceramic feed-through or whole ceramic packages with a thermally stable metallization have to be developed.

VII. CONCLUSION

The work describes optimization of the thermal stability of Lithium Niobate based SAW RFID transponders, including IDT metallization, packaging and antenna design. At 300°C a thermal stability and operability of >4000 hours was achieved, thereby outperforming standard CMOS RFID transponders by far. The given application examples of industry pilots and ongoing RFID installations demonstrate the good potential SAW RFID transponders for harsh environment.

ACKNOWLEDGMENT

The authors would like to thank the industrial partners Georgsmarienhütte GmbH, RHI AG, Identec Solutions AG and HM Energy, LLC for the successful cooperation. This project has been supported within the COMET – Competence Centers for Excellent Technologies Programme by BMVIT, BMWFJ and the federal provinces of Carinthia and Styria.

REFERENCES

[1] SOFIS Surface acoustic wave identification system, www.mobility.siemens.com/.../ds_sofis_en.pdf, Siemens AG.

[2] OIS-W Datasheet, Baumer Ident, WD.0014.EN – Edition 3 August 2003.

[3] C. S. Hartmann, P. Brown, and J. Bellamy, “Design of Global SAW RFID Tag Devices,” Proceedings of the Second International Symposium on Acoustic Wave Devices For Future Mobile Communication Systems, Chiba University, Japan, March 2004.

[4] A. Stelzer, M. Pichler, S. Schieblhofer, , and S. Schuster, “Identification of SAW ID-tags using an FSCW interrogation unit and model-based evaluation,” UFFC IEEE Transactions, 2004, pp. 1413-1420.

[5] R.Hauser, et.al. “A wireless SAW-based temperature sensor for harsh environment,” Proceedings of IEEE, Sensors 2004, pp. 860 – 863, vol.2.

[6] R. Fachberger, G. Bruckner, and A. Binder, “Durability of SAW transponders for wireless sensing in harsh environment,”Sensors, 2008 IEEE, pp. 811 – 814.

[7] A. Binder, “SAW Transponder – RFID for Extreme Conditions,” in Deploying RFID – Challenges, Solutions, and Open Issues, ISBN 978-953-307-380-4, DOI: 10.5772/17526, 2011.

[8] R. Fachberger, A. Binder, and A. Erlacher, “SAW-RFID and temperature monitoring of slide gate plates,” Sensors, 2009 IEEE, pp. 1514-1517.

[9] US patent application US 20120075113 A1.

[10] T. Aubert, O. Elmazria, J. Bardong, G. Bruckner, and B. Assouar “Is AlN/Sapphire bilayer structure an alternative to Langasite for ultra-high-temperature SAW applications ?,” Ultrasonic Symposium Proceedings, 2011 IEEE, pp. 2082-2085.

 ISBN: 978-1-61208-296-7

Advanced Metering and Data Access Infrastructures in Smart Grid Environments

Armin Veichtlbauer, Dominik Engel
Josef Ressel Center for User-Centric
Smart Grid Privacy, Security and Control
Salzburg University of Applied Sciences
Puch/Salzburg, Austria
{firstname.lastname}@en-trust.at

Fabian Knirsch, Oliver Langthaler,
Felix Moser
Research & Development, Energy Systems
Cappatec
Salzburg, Austria
{firstname.lastname}@cappatec.com

***Abstract*—For securing the provision with electric energy in smart grids, the ascertainment of energy consumption and production is inevitable in order to be able to apply appropriate control algorithms. As energy management is a highly distributed task, a data network has to exist in parallel to the power network for gathering the required data and applying control strategies by using distributed actuators, e.g., for demand response applications. For the sensing part, the introduction of Advanced Metering Infrastructures is a step ahead to the targeted meter data acquisition. Yet the definition and deployment of intelligent, but nevertheless privacy preserving, distributed control algorithms demands the collection of data from distinct sources and the application of further services like aggregation, anonymization, etc. Thus, future metering infrastructures have to provide more functionalities than those of a simple sensor network. In this paper, we discuss methods to provide such an added value infrastructure.**

***Index Terms*—Advanced Metering Infrastructures; Smart Grids; Role-based Data Access; User Control; Generic Interfaces**

I. INTRODUCTION

In order to address the challenges of a modern Smart Grid IT communication infrastructure, it is imperative to incorporate various decentralized data sources representing all kinds of different stakeholders in a Smart Grid communication environment [1]. Data items of varying precision and granularity need to be stored, processed, enriched, transported, transformed and transmitted to serve the requirements of all stakeholders, i.e., customers, energy providers and regulatory authorities.

Data from different sources in a Smart Grid environment may be retrieved in many different representations and formats. For Advanced Metering Infrastructures (AMI), there are a number of (partly proprietary) protocols available, i.e., the need for means to make these data sources interoperable is evident [2]. Furthermore, different data items may have logical relations which require further processing.

An AMI acting isolated from other services is not enough to ensure customer acceptance as well as commercial interests [3]. An all-encompassing approach for generating added value requires high level services which are able to operate across system, network and format boundaries.

While sensor networks typically operate in a limited domain only, significant enhancement of data can be achieved when being integrated in an overall network of information. However, such information does have a specific semantic relationship, e.g., a customer is assigned a meter and has a specific contract, which is not physically represented at all. There are a multitude of data sources that need to be combined to retrieve data as an additional service.

Establishing an architecture which is able to provide this functionality involves all layers of communication and information exchange. Starting from physical data exchange up to the application layer, an entire series of challenges need to be mastered.

For addressing the different layers and zones of a Smart Grid, the Smart Grid Architecture Model (SGAM) [4] is a standardized model [5] authorized by a European Commission mandate M/490 [6].

The main requirements for such an architecture are:

- Retrieving data from multiple and decentralized sources with heterogeneous interfaces and data structures
- Storing and caching data in appropriate resolutions to ensure the delivery of appropriate services
- Adding additional services for enhancement of data and information retrieval as one of the key features to ensure a sustainable metering infrastructure
- Providing data in a standardized format to allow open accessibility for client applications
- Assuring security and privacy of information and personal data by a role-based access model

 ISBN: 978-1-61208-296-7

In this paper, we propose a middleware architecture that allows for the retrieval of data from multiple sources in a standardized format and provides a scalable environment for processes and activities that use, enhance and maintain access to said data and the corresponding sources. We start with an overview of related work in Section II, followed by a description of our proposed architecture in Section III and an outline of internal data structures in Section IV. Section V presents the interfaces for concrete data providers as well as a generic interface for data users. Section VI contains information about the conducted implementation and evaluation activities.

II. Related Work

In [7], we presented a Smart Metering and data access infrastructure for a Smart Home environment. Data is retrieved, processed and provided to using applications. This architecture provides a good basis for our present research activities; yet it had two shortcomings which we had to overcome: First, the access of data sources as well as the scope of the applications were centralized; second it did not provide means for value enhancing services like data aggregation.

For Advanced Metering Infrastructures, there are several approaches dealing with different data sources [8]. As security and privacy issues do have a significant impact on design, implementation and subsequently on customer acceptance [9], such security and privacy issues have been identified and described especially concerning the intended two-way functionality of modern power grids [10].

[1] and [11] each describe a comprehensive approach dealing with communication aspects, incorporating mainly the component and communication layers of the M/490 conceptual model. In both cases, IP-based network infrastructures as well as the ensuing latency issues are discussed. Furthermore, [1] proposes an information middleware to deal issues like security and data storage. Both approaches, however, are omitting any higher level middleware functionality.

Another approach recently presented [12] introduces a Service Oriented Architecture (SOA) for a Smart Grids middleware, provided by individual web services with a centralized policy store/policy decision point. Thus, it becomes in principle possible to provide any service with any data source, yet this work focuses mainly on security and privacy issues.

Handling different types of data is one part of the challenge, finding a generic approach to storing and processing the data is a different issue altogether. An approach for adding an ontology layer for semantically sensitive retrieval is proposed by [13].

III. Architecture

Considering the requirements listed in Section I, it becomes apparent that sensor networks can only be a part of the solution, even if they are combined with classic middleware. To cover the entire set of requirements, a SOA is well suited. The advantage of a SOA over plain middleware is the possibility to define processes as a whole and not just interfaces. Such a layer provides sets of services, incorporating multiple concrete data sources. Services include simple requests expressed in a common language as well as more complex queries, where the intermediary layer provides additional value for any given data. Thus, also a hierarchical structure of abstract services according to the SGAM is made possible.

To achieve centralized data access, we designed an integrated business logic where services are tailored to data sources at a certain organizational or logical unit. This design is fully aligned with the conceptual idea of a SOA. As in [7] we defined an architecture that is designed around three layers: The infrastructure layer includes the data sources (sensors, but also database interfaces), a middleware layer (the “Core Engine”) provides services, and an application layer comprises client applications and service users.

All Core Engines are independent systems, each operating directly with their associated data sources and client applications. However, these Core Engines can communicate and interoperate with each other, as shown in Figure 1. This allows System A client applications to access - in accordance with security policies - services from System B or other connected systems. Supplemented by Single Sign On (SSO) solutions, client applications can gain transparent access to numerous systems without being directly concerned with data origin issues. Client applications themselves may act as service providers and can thus combine data from various sources.

This concept provides scalability at multiple levels: Individual cores can maintain connections to an arbitrary number of data sources. It is even possible to attach a core instance to a single meter, e.g., a Smart Meter installed in a private house. Furthermore, security policies can be defined for individual cores as well as for individual services. Policies are based on a set of rules that are applied for a certain user or a certain group. For a single core with locally assigned security rules, the instance acts both as Policy Decision Point (PDP) and Policy Enforcement Point (PEP). For a grid of connected cores, a hierarchy of trust is implemented. Assume System A is a trust providing party (“Primary

 ISBN: 978-1-61208-296-7

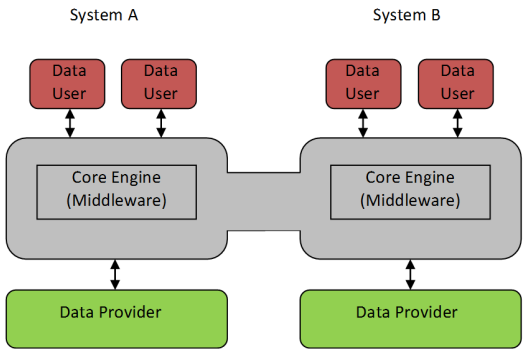


Figure 1. Platform Independent Architecture

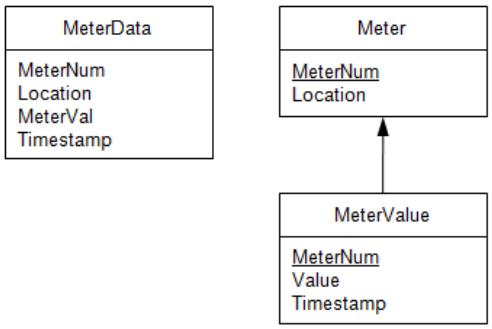


Figure 2. Example of Unnormalized and Normalized Tables

Core") for System B and System C ("Secondary Cores"). For any requests to a Secondary Core, System A acts as the PDP and the Secondary Core as the PEP.

IV. Data Structures

As data will be treated in a set oriented way, concepts of relational algebra are applied to the definition of data structures. To facilitate data processing with object oriented programming languages, certain aspects of object oriented models (such as inheritance) further influence the design of the data model. In practice, this concept is directly mapped to relational database management systems (RDBMS). Modern RDBMS (Oracle, DB2, PostgreSQL) include an extensive support of the relational set oriented model and an object oriented layer [14]. The biggest advantages of relational databases are their popularity in the industry and the standardized access via the Structured Query Language (SQL) [15], which is widely supported and implementation independent.

Another interesting approach is to use XML-based databases. It has to be distinguished between native XML databases where data is stored document centered, i.e., as it comes, and XML-based interfaces where data is stored data centered, but accessed via XML. Native XML databases seem not to be reasonable in an environment where we deal with defined data structures and not with documents. A meter will not send document oriented information, but instead a sequence of values strictly adhering to a predefined pattern. Such an approach might therefore be useful for client interfaces that retrieve data for human users, but not necessarily for automated processing facilities.

Relational databases provide the facility to map several kinds of data structures to the database [14]. Normalization levels range from completely unnormalized data structures to highly structured sets of values. The advantage of the latter is that several pieces of information can be retrieved simply by joining tables. With unnormalized tables on the other hand, it is easy to imitate big tables [14]. These databases do not keep structured table layouts but store all the information in lists assigned to a specific key. This is a powerful way to build huge hash maps for data that is always retrieved in the same manner.

Figure 2 shows an example of two different data structures both in use for our solution. They contain the same information, but for different purposes. The unnormalized table on the left is used for temporary storage of large data amounts that need to be accessed quickly and frequently. The normalized structure on the right is more suitable for analysis and storage.

Concerning internal data storage, currently all concrete implementations rely on, but are not limited to, relational databases. This model has been chosen due to freely available and powerful implementations and the ability to natively integrate such data sources into the Core Engine. All possible retrieval scenarios may be combined freely for various requests, i.e., when client level requests reference to multiple data sources, the corresponding activity might use cached data, live data and locally stored resources to service the request.

V. Interfaces

Generally, there are two types of interfaces to be considered: Data needs to be accessed by data users and by data providers, respectively. Either of these two require facilities to actively request data and facilities to be triggered by the middleware to retrieve data.

A. Data Providers

Although a number of interfaces to data sources will need to be incorporated in a Smart Grid / Smart Metering

 ISBN: 978-1-61208-296-7

infrastructure, within the scope of this work we focus on test data provided by our company partner Salzburg AG. This comprises the following data sources:

- SAP IS-U: An SAP system providing customer and technical data. Concrete data sets include customer name, contract information and meters assigned to a single customer. Data is retrieved through SAP function modules. Function arguments are a business partner ID. This model is commonly represented as the *SAP IS-U House* [16].
- AMIS: A Meter Data Management system providing actual meter data [17]. This database archives actual data from Smart Meters representing customer usage, meter states and latency values. Data is either retrieved through a web service or directly via native database interfaces. Web service arguments are a meter number and a period of time, returning customer consumption.
- SMIS: A database providing specific meter data for gas and district heating. Data is accessed directly from a relational database through an XML interface provided by the local system.

All these interface provide data in different formats and with different protocols.

B. Data Users

Designing an application interface for data users particularly involves the following functionality:

- An open standard for mark-up that can easily be implemented by the different kinds of client applications
- A protocol for authentication/authorization and persistent operations
- A data exchange protocol that can be defined upon the used open standard

Furthermore, any connections to the system should converge to a single interface definition in order to keep homogeneity throughout all aspects of client accessibility. To meet our previously defined requirements, XML has been chosen due to the vast number of freely available implementations for any client application [18] and its ability to be easily transformed from one scheme to another [19]. The concrete interface has been created by defining XML-based data structures to handle all kinds of data types (such as numbers, strings, timestamps, binary objects) and allowing scalable representations from scalar values to entire sets containing multiple rows. This interface therefore implements a mapping of the concepts used for data structures to actual data representation for external components.

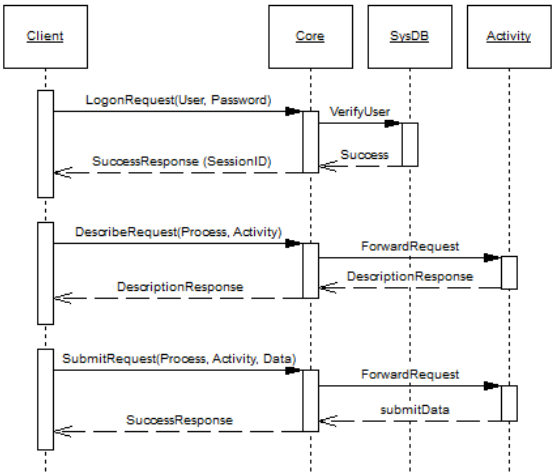


Figure 3. Data Retrieval Process

For our solution we used R×XML [20] as a single interface to connect to external client applications, other instances of the Core Engine and to all data sources. R×XML defines a set of XML schemes that has been designed to provide both an efficient and comprehensive representation of data for transfer or storage and a protocol to exchange basic information about access rules and persistency. In order to support various web-based protocols, such as SOAP or REST, a "SOAP Intermediary" component is applied as a converter to R×XML. This allows multiple cases in which the tasks of the Intermediary can vary from simply transforming data from clients to the Core Engine to providing additional value, such as generating graphical representations for web applications.

This additional interface is not integrated directly into the Core Engine but builds upon the existing R×XML interface. It is designed as a separate and independent service. Clients can communicate with this server by sending web requests via the SOAP protocol. Incoming messages are parsed and their SOAP body is converted to R×XML. Subsequently, the SOAP Intermediary forwards the data to the Core Engine for further processing. The Core Engine in turn regards the SOAP Intermediary as a user with specific authorizations.

In accordance with the aforementioned concept of PDP/PEP, security is governed by the Core Engines solely. A trusted provider (i.e., the owner of the specific Core instance) can label certain processes and activities with access privileges for specific users and roles. When clients log on, they provide a username/password combination to the Core Engine and are assigned a Session ID. Based on this Session ID (a random n bit hex

 ISBN: 978-1-61208-296-7

string), further requests are handled. Every Session ID can be mapped directly to a certain user and therefore to a certain role and a set of access rules for that role. Figure 3 shows a sequence diagram illustrating the data retrieval process. In this scenario, a data user (client) logs on, retrieves an interface description and subsequently retrieves data. For securing the transmission of requested data HTTPS [21] is used. As the current implementation is enforcing end-to-end security, no additional encryption is required. However, for specific clients / data sources, data might be custom encrypted and then transferred over other protocols.

VI. Implementation Aspects

For software development, we mapped the architectural outcomes to concrete technologies and design patterns. The general architecture has been developed strictly adhering to the Model-View-Controller (MVC) layout. Data sources, business logic and client representation are conceptually linked via a set of defined interfaces. MVC is a direct mapping of the proposed three-layer architecture [22]. Figure 4 shows the architecture of the implemented prototype.

For process controlling, activity management and role-based user management, the *Cappatec Core Engine 1.x* is used. This Core Engine is enterprise software, developed to run all kinds of applications and services, including business processes and activities, on a common platform. It has been determined that within the scope of the project all the specified requirements and all the use cases developed for an advanced metering and data access infrastructure can be represented as activities on this platform.

The fundamental concept of this approach is the definition of individual modules, which incorporate the business logic by handling certain tasks, such as retrieving and/or aggregating data. These modules may be developed completely independently of each other and then executed on a central Core Engine, which handles runtime behavior and governs security issues.

As an environment for development and testing, a VMware vSphere 5 Update 1 Hypervisor running two instances of CentOS 6.3 as well as one instance of Windows Server 2012 has been set up. This way, easy migration to other, more powerful hardware platforms could be ensured. The first instance of CentOS was used as a host for the Cappatec Core Engine, while the second instance was running PostgreSQL 9.2.3, containing the system and cache databases. Windows Server 2012 was used as host for the SOAP Intermediary.

Development of the Cappatec Core Engine has been performed using Eclipse Juno as IDE and Java 7 as programming language. Furthermore, a lightweight client written in C# and JavaScript, which provides generic access to the Core Engine for debugging purposes, has been developed using Visual Studio 2010. The SOAP Intermediary utilizes Apache Tomcat 4.1.3 and Axis 1.4 and has been written in Java 7.

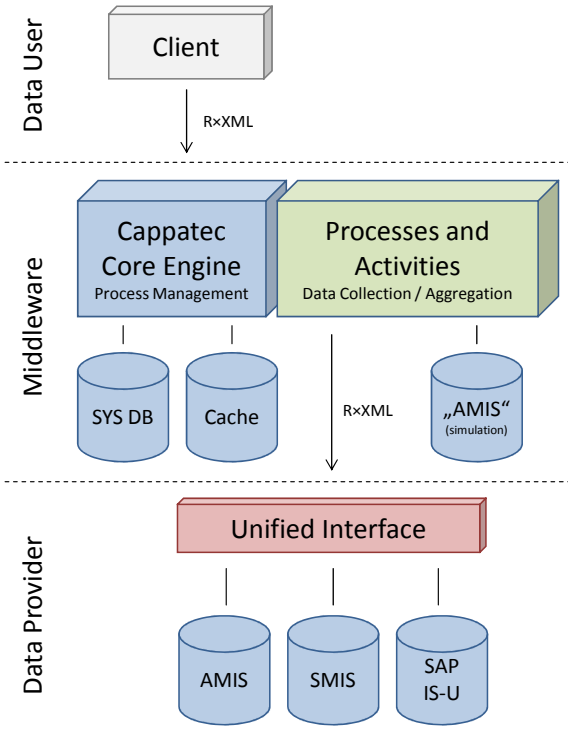


Figure 4. Platform Specific Architecture

VII. Validation and Conclusion

The network setup during development and testing has been kept as simple as possible. All three of the mentioned virtual images have been placed within one local subnet. To achieve R×XML convergence across all aspects of data retrieval (e.g., contract information, meter data), an instance of Inubit 6.1 [23], which connects to all relevant subsystems of the energy provider and acts as an R×XML converter, has been utilized. Conversion is performed by means of XSLT, which is supported natively by this platform.

For evaluating the prototypical implementation, we considered two data sources:

- First, we processed real, but anonymized measurement data from Salzburg AG, connected via Inubit, in order to validate the accessibility of real measurement data.

 ISBN: 978-1-61208-296-7

- Second, we used simulated data in an internal AMIS data base to make it easier to compare multiple scenarios and error conditions, as well as for scalability tests.

Both data sources showed the basic functionality of our solution, i.e., the access to measurement data and the provision to respective applications. Yet the actual strength of the approach is the integration of additional services like data aggregation; here, further validation efforts will be necessary. For instance, our research group has particular interest in the combination of measurement data from the AMIS system and the business data from the SAP IS-U [16]. Also, further tests will be necessary in order to compare our solution with existing ones, e.g., in terms of performance issues.

VIII. Further Work

Besides exhaustive validation, further research might be carried out especially in the domain of security and privacy. The authentication of data user, data providers and distributed Core instances can be performed by certificates. Concerning the architecture itself, the use of a completely generic and semantic data model is considered. Such a data model, however, should integrate seamlessly in the entire architecture and must therefore provide an additional set of semantic interfaces to data providers and users.

Acknowledgments

The work described in this paper was conducted in the "Josef Ressel Center for User-Centric Smart Grid Privacy, Security and Control", which is supported by the Austrian Federal Ministry of Ecomomy, Family and Youth (BMWFJ) and the Austrian National Foundation for Research, Technology and Development. The authors would also like to thank René Blaschke from the company partner Salzburg AG for providing extensive test data.

References

[1] Y. Kim, M. Thottan, V. Kolesnikov, and W. Lee, "A Secure Dezentralized Data-Centric Information Infrastructure for Smart Grid," *IEEE Communications Magazine*, vol. Energy Efficiency in Communications, pp. 58–65, Nov. 2010.

[2] Z. Fan, P. Kulkarni, S. Gormus, C. Efthymiou, G. Kalogridis, M. Sooriyabandara, Z. Zhu, S. Lambotharan, and W. Chin, "Smart grid communications: Overview of research challenges, solutions, and standardization activities," *IEEE Communications Surveys & Tutorials*, vol. PP Issue:99, pp. 1–18, 2012.

[3] C. Müller-Elschner, "Die Rolle von Informations- und Kommunikationstechnologie beim Smart Metering," in *Smart Metering – Technologische, wirtschaftliche und juristische Aspekte des Smart Metering*, 2nd ed., C. Köhler-Schute, Ed. KS-Energy-Verlag, 2010, pp. 87–95, in German.

[4] *Smart Grid Reference Architecture*, CEN/Cenelec/ETSI Smart Grid Coordination Group Std., Nov. 2012.

[5] *Framework Document*, CEN/Cenelec/ETSI Smart Grid Coordination Group Std., Nov. 2012.

[6] M. S. Jimenez, *Smart Grid Mandate*, European Commission Directorate-General for Energy Std., 2012.

[7] A. Veichtlbauer, T. Pfeiffenberger, and U. Schrittesser, "Generic control architecture for heterogeneous building automation applications," in *Proceedings of the 6th International Conference on Sensor Technologies and Applications (SensorComm 2012)*, Rome, August 2012, pp. 148–153.

[8] ITU-T, "Applications of ITU-T G.9960, ITU-T G.9961 transceivers for Smart Grid applications: Advanced metering infrastructure, energy management in the home and electric vehicles 2010," ITU-T technical paper, Series G: Transmission Systems and Media, Digital Systems and Networks.

[9] D. Engel, "Wavelet-based load profile representation for smart meter privacy," in *Proceedings of the Fourth IEEE Conference Innovative Smart Grid Technologies (ISGT'13)*, Washington, D.C., USA, Feb. 2013, pp. 1–6.

[10] A. Barenghi and G. Pelosi, "Security and privacy in smart grid infrastructures," in *Proc. 22nd Int Database and Expert Systems Applications (DEXA) Workshop*, 2011, pp. 102–108.

[11] V. Sood, D. Fischer, J. Eklund, and T. Brown, "Developing a communication infrastructure for the Smart Grid," in *Electrical Power Energy Conference (EPEC), 2009 IEEE*, Montreal, QC, Canada, Oct. 2009, pp. 1 –7.

[12] M. Jung, T. Hofer, S. Döbelt, G. Kienesberger, F. Judex, and W. Kastner, "Access control for a Smart Grid SOA," in *Proceedings of the 7th IEEE Conference for Internet Technology and Secured Transactions*, London, UK, Dec. 2012, pp. 281–287.

[13] S. Rohjans, "(S2)In – Semantic Service Integration for Smart Grids," Ph.D. dissertation, Carl von Ossietzky University, Oldenburg, Sep. 2012.

[14] G. Vossen, *Datenmodelle, Datenbanksprachen und Datenbankmanagementsysteme*, 5th ed. Oldenbourg Verlag, 2008, in German.

[15] ISO/IEC, *ISO/IEC 9075-1:2011 – Information technology – Database languages – SQL*, International Organization for Standardization and the International Electrotechnical Commission Std., 2011.

[16] Ramamoorthy, K. (2012) SAP for Utilities: SAP IS-U data model - Business and Technical Master Data. Accessed: 2013-06-06. [Online]. Available: http://scn.sap.com/community/utilities/blog/2012/03/26/sap-is-u - data-model–business-and-technical-master-data

[17] Siemens AG. (2013, Mar.) Information mit System – Das automatisierte Verbrauchsdatenerfassungs- und Informationssystem AMIS. In German. Accessed: 2013-06-06. [Online]. Available: http://www.siemens.com/sustainability/pool/de/umweltportfolio/produkte-loesungen/energieuebertragung-energieverteilung/amis-broschuere-de.pdf

[18] H. Vonhoegen, *Einstieg in XML – Grundlagen, Praxis, Referenz*, 5th ed. Galileo Computing, 2009, in German.

[19] J. Clark. (1999) XSL Transformations (XSLT) Version 1.0 – W3C Recommendation. Accessed: 2013-06-06. [Online]. Available: http://www.w3.org/TR/1999/REC-xslt-19991116

[20] F. Knirsch and O. Langthaler. (2013, Feb.) Cappatec Core Engine RxXML Specification. Accessed: 2013-06-06. [Online]. Available: http://www.cappatec.com/home/res/Cappatec_RxXML_02_13.pdf

[21] E. Rescorla, "HTTP Over TLS," RFC 2818 (Informational), Internet Engineering Task Force, May 2000, accessed: 2013-06-06. [Online]. Available: http://www.ietf.org/rfc/rfc2818.txt

[22] I. Sommerville, *Software Engineering*, 6th ed. Pearson Education, 2001.

[23] Bosch Software Innovations. (2013, Feb.) Inubit for Business Process Management. Accessed: 2013-06-06. [Online]. Available: http://www.bosch-si.com/technology/business-process-management-bpm/business-process-management.html

 ISBN: 978-1-61208-296-7

Anchor-free Localization in Wireless Lamp Networks using Superimposed RSSI Measurements

A. Caracas, T. Eirich, T. Kramp, M. Oestreicher
IBM Research - Zurich
(xan,eir,thk,oes)@zurich.ibm.com

M. Hoffman, C. Gargiulo, G. Soros
ETH - Zurich
(moritzho,gargiulo,gabor.soros)@ethz.ch

***Abstract*—Localization is a crucial aspect in providing context-aware services and applications. Existing approaches often require a fixed infrastructure or specialized hardware, providing varying degrees of accuracy. For certain application scenarios, high accuracy is not required, if the objects to localize are fixed or slowly moving. For such relatively static scenarios, we propose a novel RSSI-based localization technique using multiple antennas and multiple channels in the 2.4 GHz band. As a concrete application, we present a wireless lighting control system for office lamps in an Internet of Things (IoT) scenario. The lamps participate in a wireless sensor network (WSN) and cooperatively adjust the ambient light based on the self-determined topology correlated with occupancy information. Our experimental results show that we are able to achieve an accuracy of 2 m in 93% of distance estimates with cost-effective IoT technology, improving existing results by 1 m.**

***Keywords*—RSSI-based localization, multi-antenna, multi-channel, wireless lighting**

I. Introduction

Localization is a crucial aspect for context-aware applications and services. We take for granted and rely on the suggestions given by our navigation systems when driving to a new destination, or, we check the 15-minutes updates of the rain radar before starting the bike journey home. The most prominent outdoor localization system is the satellite-based global positioning system (GPS). However, GPS is not able to provide accurate information or no information at all in certain scenarios. These scenarios occur when the line of sight to GPS satellites is obstructed, in particular, for indoor environments [1] as well as outdoor, in dense forest areas [2], [3] or mountainous regions. In such cases, wireless sensor networks (WSNs) represent a viable, and often the only possible alternative for enabling positioning services [4] in local areas.

In this paper, we focus on improving radio-based localization using the standard received signal strength indicator (RSSI), which is a measurement readily available on off-the-shelf WSNs. One important criterion for our localization system is to operate without any additional infrastructure, e.g. setup and configuration, no fixed beacons, and no extra cabling.

To test the feasibility and accuracy of our proposed localization approach, we concentrate on a concrete application scenario: a wireless lighting control system. We enable lamps, positioned in large, open-office environments, to autonomously determine their location and topology information. Based on this information, and combined with motion detection information, the lamps cooperate in adapting and creating ergonomic ambient lighting conditions.

This paper is structured as follows. In Section II, we discuss related wireless localization systems and how our localization technique advances the state of the art. Following, in Section III, we describe the details of our proposed localization technique. As a concrete application, in Section IV, we focus on an example scenario for a wireless lighting control system. In Section V, we present our experiments and evaluate the results of our localization technique in the context of the lighting control scenario. Finally, in Section VI, we conclude the paper and provide an outlook for future work.

II. Related Work

Localization is an extensively researched topic. Different technologies exist to determine the location of objects in a global or relative frame of reference using various signals (e.g., radio, light, temperature or sound). Due to their extended range, radio-based localization (RBL) systems can be applied to a wide range of environments [4].

In an RBL system, an estimate for the distance between two objects can be obtained based on several techniques such as time-of-flight (ToF), angle of arrival (AoA), or time-difference of arrival (TDoA), e.g. between an acoustic signal and a radio signal. However, such techniques require highly-specialized and time-synchronized hardware [4]. An cost-effective alternative is the received signal strength indicator (RSSI), measured automatically by modern radio chips. This indicator can be related to distance using different path-loss models [5]. However, multipath fading effects significantly affect the accuracy [1] of any RBL system irrespective of the technique for distance estimation.

The best-known RBL example is the global positioning system (GPS). Most RBL approaches, such as GPS, work well for outside environments, far from obstacles or disturbances which may affect the signal propagation. At present, these RBL systems typically utilize a costly infrastructure, based on fixed beacons (in the case of GPS, these beacons are sent by mobile satellites orbiting the Earth) with known positions. For indoor settings, Ubisense [6] provides a commercial RBL solution, based on a fixed beacon infrastructure and highly specialized hardware tags. The tags are able to measure both the AoA and TDoA of beacons signals, which are broadcasted by the fixed infrastructure. Typically such systems operate in the 6-8 GHz (ultra-wide) band as well as the 2.4 GHz band.

 ISBN: 978-1-61208-296-7

In contrast to specialized solutions, the standardized IEEE 802.15.4 radio technology in the 2.4 GHz and 800-900 MHz frequency bands, used by modern WSNs, promises a relatively cheap and off-the-shelf solution for realizing location-based applications and services. This technology is pervasive and applicable to both indoor and outdoor Internet of Things (IoT) scenarios. Because RSSI measurements come at no additional costs, they are performed automatically by the underlying radio hardware of WSNs, we focus on such techniques for eliciting the distance estimates in a cost-effective manner.

Previous work on RSSI-based localization failed to deliver results which were accurate enough for context-aware applications. The best result [7] using RSSI-based localization showed, with off-the-shelf WSN components, an absolute error of less than 3 m (with a probability of 95%). In this case, the application scenario involved a beacon-based infrastructure for improving the tracking of a moving shopping cart in combination with odometric measurements from an inertial measuring unit (IMU). More recent work [8] in this space suggests that using multiple channels can improve the accuracy of RSSI-based localization. However, the authors present no values for the error of the distance estimation.

Most localization approaches focus on tracking moving targets such as controlling cattle in agriculture [9], correlating monitoring information in logistics applications[10], locating shopping carts [7] for marketing purposes, and ensuring rapid response times when elderly persons in assisted living environments [1] are involved. Moreover, available solutions often require a fixed infrastructure, which must be minutely calibrated [6], [1] for best results. However, interesting and business-relevant application scenarios can emerge where relatively static objects (e.g. lamps, computing assets, or shared lab equipment) are able to determine their location without the need of a fixed infrastructure. The self-determined location can be subsequently used as the basis for localizing moving targets.

To summarize, our approach focuses on providing a pervasive, cost-effective, and self-configurable localization solution. Our main contribution addresses the multipath fading by combing RSSI information from multiple frequencies as well as multiple antennas for the same physical location. While we focus on a static scenarios our approach can be extended for tracking moving targets, once the self-configured infrastructure is in place.

III. A Novel Localization Approach

To counteract the detrimental effects of multipath fading, we propose leveraging the different frequencies (also called channels) combined with spacially-separated observations for a given physical position. Such observations are achieved by using multiple antennas, at least half a wavelength apart. For practical purposes, to emulate multiple antennas, we use multiple motes placed next to each other in the same location. To enable a commercial solution, multiple antennas can be controlled by the same radio chip and integrated into a single-mote design.

A. Rationale for Multiple Channels and Multiple Antennas

In an ideal case, i.e., without multipath fading, the same RSSI measurements will be reported for all frequencies. (or channels) and for all antennas. By varying the channel, the wave-length is varied, which in turn changes the propagation paths of the indirect signals, because of different reflection patterns. For the direct radio signal, based on Friis's equation [5], the received power is proportional with the square of the wavelength. However, it can be easily shown that for the different wavelengths (of ≈12.5 cm) in the 24 GHz band, the received signal strength would vary by less than 1% at distances of more than 1 m. In other words, some propagation paths will not occur, while others may be induced. The same effect is introduced by placing multiple, spacially-separated antennas in the same location. Intuitively, overlapping RSSI measurements, on different channels, and observed by different antennas, are a very strong indicator for the presence of a component for the correct RSSI value. The RSSI candidate value can then be computed as a linear combination of these strong components.

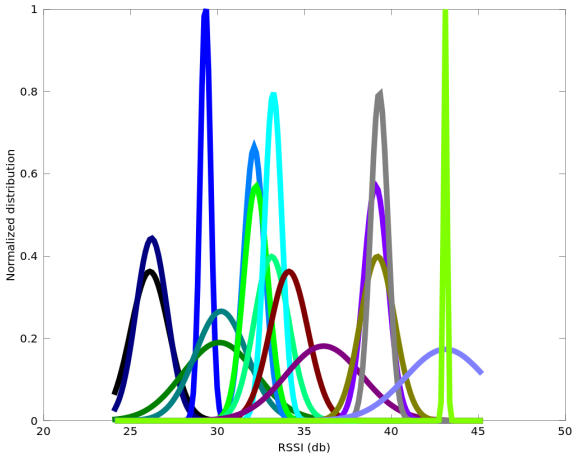


Fig. 1. The distribution of RSSI measurements for individual channels as measured by an observer mote for a given observed mote. When considering the opposite observation direction the RSSI measurements typically follow a similar distribution.

Using multiple channels and multiple antennas for performing RSSI measurements introduces a new challenge: the different channels report different, albeit relatively stable, RSSI measurements. Fig. 1 shows, for each channel, the distribution of RSSI measurements of the messages sent by the observed mote and received by the observer mote. In our experiments, the difference in the stable RSSI measurements for different channels and different antennas, for the same physical location, were as much as 25 db. The challenge in this case is to select an RSSI value which best represents the distance between the observer and observed mote.

Literature [8] suggest that averaging RSSI measurements across all channels can improve the RSSI-based estimations. However, such a straight-forward approach is biased and extremely sensitive to noise.

B. Selecting the Best RSSI Candidate

Our proposed algorithm for selecting the best RSSI candidate operates in three stages. Because a pair of motes (or

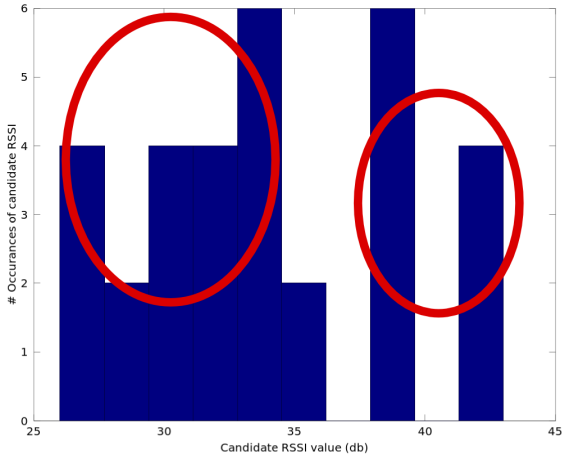


Fig. 2. Distribution of RSSI candidate values for a given pair of motes. The x-axis represents the RSSI candidates, whereas the y-axis shows how many times the RSSI value was selected as a candidate value, for both observation directions and across all channels.

group of motes) corresponds to a single distance, both RSSI observation directions are considered for each distance pair.

In the first stage, the algorithm creates a distribution (histogram) of RSSI candidates. For each channel, and for each observation direction, a candidate is selected based on the most stable RSSI value from the given measurements. Selecting the stable value instead of the mean (which is sensitive to noise) has a significant effect on the final result. Note that in noisy situations, for a given channel and pair combination, it is possible that no selection is made. Effectively, only selections with a minimum confidence (given by the coverage ratio of the stable value) are kept.

Fig. 2 shows such an example distribution of the most stable RSSI candidates for a given pair of motes, where both observation directions are combined. The x-axis shows the candidate RSSI values, whereas, the y-axis shows the number of occurrences (across all channels and both observation directions). For this example, the RSSI candidate value of 39 db was selected 6 times (on 6 channels), based on the measurements performed in both observation directions.

Because we operate on all 16 channels, we obtain a maximum of 32 RSSI selections for each distance pair. In the general case, when grouping motes (treated as multiple antennas) for each distance pair we obtain a maximum of $32 \cdot k^2$ candidate RSSI values, where k is the number of motes (or antennas) in a group.

In the second stage, we filter from the histogram all RSSI values with a low number of occurrences. A low occurrence means there is no correlation with measurements from other channels. The filtering step also separates clusters of RSSI values which are at least 5 db units away from each other. We consider these clusters (shown as ellipses in Fig. 2) to be the strong components for the final RSSI value.

Clusters (marked by ellipses) of most frequent candidates indicate strong components for the actual RSSI value. Note that in comparison to Fig. 1, for a given pair, both observation directions are used. The linear combination of clusters gives an RSSI value which best represents the given pair in terms of the estimated distance. For each cluster c, we then compute the cluster average r_c. Finally, we compute the best RSSI candidate R, which is later used for estimating the distance, as the weighted average of clusters

$$R = \sum_{i=1}^{K} w_c r_c \quad (1)$$

where K is the number of clusters and w_c is the weight for each cluster, obtained as the sum of candidates in each cluster.

In the final stage, we estimate the distance $d(R)$ using the exponential decay path-loss model, which has been widely validated by experimental data [10], [11], [12]. Using this model, the distance estimate can be computed as:

$$d(R) = 10^{\frac{G-R}{10f}} \quad (2)$$

where G is the gain constant, and f is the damping factor.

The damping factor f is specific to a certain environment, whereas, the gain constant G is specific to a certain mote hardware. These parameters can be determined experimentally by applying a least-square means optimization on the measurements with known distances. Alternatively, the parameters can be obtained from previous experimental results [12].

C. A TDMA-based RSSI-collection protocol

To rapidly construct different topologies and for performing RSSI measurements on large areas (requiring multi-hop communication) over extended periods of time, we designed and implemented a custom TDMA-based protocol. We are thus able to gather the individual RSSI observations of each mote in the network. A TDMA-based approach allows the motes in the network to save power because measurements can be performed only when messages are expected. Thus, longer periods of operation are possible with the initial set of batteries. Our protocol implementation uses the IBM Mote Runner platform [13] for WSNs.

Fig. 3 shows an overview of the time slots in the protocol. The protocol uses beacon messages (B) for several purposes: network synchronization, RSSI measurements, and encoding (the computed) topology information. At the beginning of each slot, every mote listens for beacons for a short period of time. This process ensures that the network remains synchronized. Beacons also trigger the initial network formation. Shortly after the beacon, a child mote can request a parent mote for an association (A), which determines the routing tree and occurs in the network formation step.

Once the network is formed (i.e., the routing tree and child parent relations between motes are established), the RSSI values for all received beacon messages are recorded. Beacons are sent by all motes at the beginning of slots of 200 ms. Slots are pre-allocated based on the mote unique identifiers. Based on a round-robin schedule, beacon messages are sent on the full spectrum of available channels (16 channels from 0x0B to 0x21) in the 2.4 GHz ISM band of the IEEE 802.15.4 standard (commonly available for low-power WSN). For each superframe (which acts as a container for all slots), the same

 ISBN: 978-1-61208-296-7

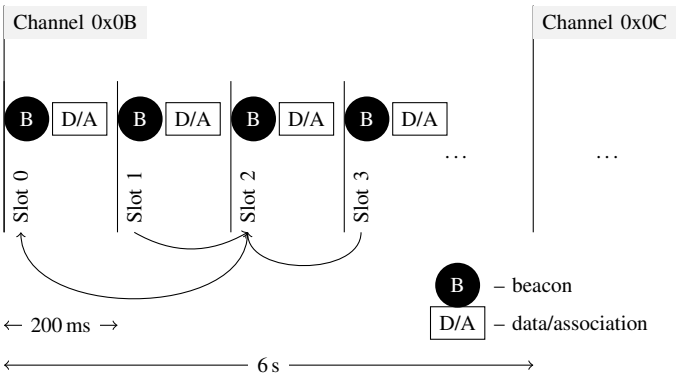


Fig. 3. Overview of slots in the TDMA-based protocol used for collecting RSSI measurements. RSSI values are measured on beacon (B) messages. The measurements are forwarded using data (D) messages, whereas association (A) messages are used to form the routing tree. The arcs show the routing tree between motes to which the slots are assigned to. In this example, the motes in slot 1 and 3 are children of mote 2, which in turn is a child of the edge mote 0. For this example, the slots greater than 4 are yet to be used.

channel is used. In our current implementation a superframe last 6 s, supporting a maximum of 30 slots (or motes). Both the slot duration and number of slots are configurable parameters and can be scaled to accommodate for larger networks.

The remaining time (after the beacon) in a slot is used by child motes to communicate the recorded data (D) to the respective parent. Effectively, motes without children only need to record beacons from other motes and sleep the remaining time. Parents append all the data collected from children to their own data and forward the entire set of measurements towards the edge mote.

Finally, The edge mote is attached to a gateway PC, which saves the measurements locally. The RSSI measurements are then analyzed periodically and transformed into distance estimates.

D. Topology from Distances

Computing a physical topology from distance estimates is a well-studied problem and several generic algorithms exists [4], [14] to address this issue. In our experiments, computing the topology using an iterative trilateration [4] approach deemed sufficient for reconstructing a recognizable map for our physical deployments. For larger deployments, which must scale with the number of nodes and limit the cumulative trilateration errors, algorithms derived from the classical multi-dimensional scaling (MDS) method, such as MDS-MAP [15] are better suited [14].

IV. A Wireless Lighting Control System

To demonstrate the feasibility and effectiveness of our localization approach, we are developing a wireless lighting system for open offices. The goal is to construct a system that can autonomously determine the topology of lamps. Based on the self-determined topology combined with motion detector and light sensors, the lamps collaboratively control their outputs in order to create an ergonomic working environment.

The advantages of such an autonomous lighting system is that it requires no time-consuming configuration step compared to existing, wired, central-control systems. Moreover, our system relies on cost-effective technology readily available in modern WSNs and there are no cost associated with cabling. Effectively, such a wireless lighting control system can be installed not only during the construction phase of new buildings but it can also retrofit older (historical) buildings without requiring structural modifications.

Another important requirement for such a wireless system is to allow for convergence times on the order of several hours. Intuitively, best results are achieved outside of normal office hours, i.e., at night time, when (most) humans (and their personal wireless devices) are not present. To a certain extent, the localization algorithm should be robust against moving humans and the noise present during normal office operations. To achieve such robustness, the algorithm requires a longer convergence time, on the order of 4-6 hours. The only assumption is that the objects to be localized, the lamps in our case, are static for the duration of the topology computation.

Effectively, using our system, the lamps only need to be placed at their intended location. Subsequently, they construct a wireless network for bi-directional communication fully autonomously. The wireless network serves as the backbone for measuring the radio signal strength, used to derive the topology of the WSN motes, which are attached to the physical lamps. The topology together with information about motion detection and actual sensed light represent the input which determines the output of the individual lamps. Thus, a finer and context-aware control over ambient lighting can be achieved.

Studies have shown that the lighting conditions and productivity in office environments are correlated [16]. Proper control of lighting and blinds is a crucial factor in reducing the operating costs of buildings [17]. Lighting can also induce psychological effects. For example, anxiety is a natural reaction in a large office space, where the individual working area is lit but the boundaries are entirely dark. In such cases, anxiety can be reduced, and the comfort can be increased, if the perimeter and the walls of the office-space are lit.

A. System Architecture

Our wireless lighting solution is structured in three major components as shown in Fig. 4. The office lamps themselves represent the core components. Following the IoT paradigm, each lamp is augmented with an off-the-shelf mote to allow the lamp to participate in a wireless network, which is responsible for running the protocol described in Section III-C for collecting RSSI measurements.

In a commercial setting, the lamp design will be fully integrated with the components of a WSN mote. In other words, the antenna for wireless communication would be indistinguishable from the lamp design, and the MCU and radio chip will be hidden from view. A further design factor may consider using power directly from the lamp supply for powering the WSN mote, instead of batteries currently employed by our development WSN platform.

The second major component of our system is a gateway composed of an embedded PC, with more compute power

 ISBN: 978-1-61208-296-7

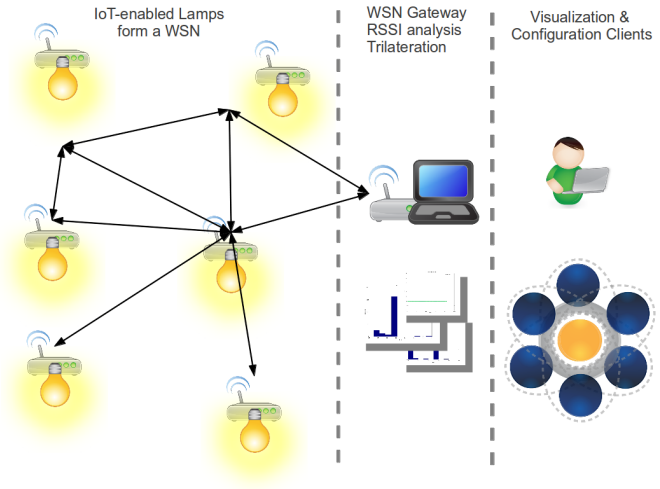


Fig. 4. Architecture overview for the wireless lighting control system. The IoT-augmented lamps form a WSN for collecting RSSI measurements used for automatic localization.

available than the motes attached to each lamp, and an edge mote for connectivity with the WSN. In our current architecture, the gateway is responsible for performing the localization computations and communicating the respective results back into the WSN. In a commercial setting, the localization algorithm can be adapted to run directly on the wireless motes. The gateway runs an embedded Linux distribution and creates a public WiFi (IEEE 802.11) access point to which the visualization client can connect to.

Finally, the visualization component displays the computed topology and the state of the WSN. In addition, the user interface is also intended for setting desired ambient parameters for creating a pleasant and ergonomic working environment. For example, it should be possible to enable a light fading effect towards the perimeter, or setting a brighter lighting in the center of the office space.

V. System Experiments and Evaluation Results

To test the feasibility and accuracy of our localization solution, we conducted several experiments both indoors and outdoors (see Fig. 5). In both cases motes are placed on wooden posts. To emulate multiple antennas multiple motes were placed in groups at the same physical location.

For our experiments, we used standard, off-the-shelf wireless motes, which were deployed in homogeneous networks. To study the effects of different hardware and antennas, we used two platforms: Memsic IRIS motes, with a quarter wavelength dipole antenna, and Atmel AVRRAVEN motes, with a loop printed antenna. Both hardware platforms are based on the same ATMEGA1281 micro-controller operating the same RF230 radio chip, and both run our RSSI-collection protocol on top of the IBM Mote Runner OS [13].

The initial step in our evaluation is determining the parameters for the path-loss model: the damping factor f and the gain constant G. Our measurements confirmed previous experimental results for the damping factor, i.e. $f = 1.8$ for the indoor (corridor) settings and $f = 2.5$ for the outdoor

(a) (b)

Fig. 5. Indoor (a) deployments using IRIS motes and outdoor (b) deployments using AVRRAVEN motes.

setting. Note that the damping factor applies to both the IRIS and the AVRRAVEN hardware. By contrast, to accommodate for the difference in the antennas, based on the hardware specifications, we used $G = 55$ for the AVRRAVEN motes and $G = 45$ for the IRIS motes respectively.

Our deployments used a maximum of 24 motes, which were deployed in different topologies: linear as well as rectangular meshes. To test the effects of spacial separation of antennas we deployed groups of one (1X), two (2X), three, and respectively four motes in the same physical position. The maximum covered area for the outdoors deployments was a field of 200 m^2, with a maximum distance of 20 m between motes. For the indoor deployments, in offices and corridors, we covered a maximum area of 50 m^2, with a maximum distance of 15 m between motes. For both the indoor and outdoor deployments, the motes were placed on wooden (70 cm high) posts to avoid high path fading effects due to proximity to the ground. Measurements were collected over periods of 24 h.

Fig. 6 compares the cumulative relative error for different experimental settings. The y-axis shows the probability that the absolute error of the estimate is below a value of x meters as represented by the x-axis. Our results show that using multiple channels significantly improves the distance estimates when compared to using a single channel only. This result is applicable for both single motes as well as groups of two motes for the same physical location. The results also show that a further improvement of distance estimation is achieved by using two antennas for the same physical location when compared to single antennas. Groups of two antennas (motes) using 16 channels for RSSI measurements achieve a maximum error of 2 m in 93% of distance estimates. Note that simply averaging all RSSI values on all channels increased the error by a few meters. By contrast, single motes using one channel for RSSI measurements, achieve a maximum error of 4.3 m in 90% of the cases.

Our results show that the best estimates are achieved by using multiple channels together with groups of 2 motes placed at the same physical location. In this case, for distances of up to 12 m, we are able to provide distance estimates, which have a maximum error of less than 2 m for 93% of the estimates. The maximum error for all estimates is 2.7 m. Our results reduce the absolute error of distance estimates by more than 1 m compared to previously reported results [7]. For our wireless lighting scenario, where lamps are positioned at least 8 m apart, the absolute errors are well within bounds.

Our experiments also showed that more than two (even

 ISBN: 978-1-61208-296-7

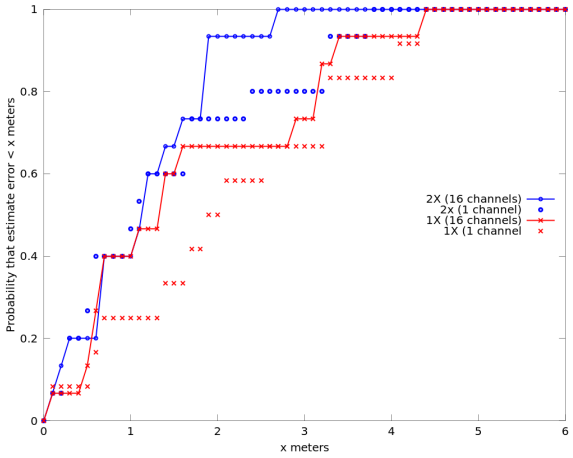


Fig. 6. Comparison between the cumulative relative error for distance estimates for pairwise groups of 1 mote (1X) and 2 motes (2X) using all 16 channels or only a single channel.

though passive) antennas can act detrimentally. In some cases, for three and four antennas, the distance estimates are worse compared to using a single antenna. This represents an interesting aspect for optimizing the antenna design, which is out of the scope of this paper.

VI. Conclusion and Future Work

In this paper, we presented a novel RSSI-based localization approach, which combines measurements from multiple radio channels and observed by multiple antennas. Our results show that the pervasive, off-the-shelf, and cost-effective WSN technology is able to provide an accuracy of 2m for 93% of estimates for a range of up to 12 m when using two antennas and 16 channels. Such accuracy levels are sufficient for many classes of indoor and outdoor applications.

As a showcase, we presented a wireless lighting control system, currently under development, which allows office lamps to autonomously determine their topology, used to control the lighting. Our localization approach can also be applied to other scenarios, e.g., to eliminate time-consuming, manual inventories for locating computing assets or shared lab equipment. Furthermore, our approach works without the installation and configuration steps required by a fixed beacon infrastructure. Once in place, such a relatively static and self-localized infrastructure can be used to localize mobile objects for both indoor and outdoor scenarios.

For a commercially-viable solution, we envision that the trilateration algorithm will be distributed in the WSN itself, effectively eliminating the need for a gateway. Furthermore, we foresee that the visualization client (e.g., tablet PC or smartphone) can directly connect to any mote in the WSN. This requires either that the smartphones are equipped with radios for the IEEE 802.15.4 standard or that the WSN motes are able to respond using near filed communication (NFC), Bluetooth, or WiFi which represent the norm for wireless communication in the smartphone space.

References

[1] M. A. Stelios, A. D. Nick, M. T. Effie, K. M. Dimitris, and S. C. A. Thomopoulos, “An indoor localization platform for ambient assisted living using UWB,” in *Proceedings of Advances in Mobile Computing and Multimedia (MoMM)*. ACM, 2008, pp. 178–182.

[2] C. Bo, D. Ren, S. Tang, X.-Y. Li, X. Mao, Q. Huang, L. Mo, Z. Jiang, Y. Sun, and Y. Liu, “Locating sensors in the forest: A case study in GreenOrbs,” in *Proceedings of INFOCOM*. IEEE, 2012, pp. 1026 –1034.

[3] W. Xi, Y. He, Y. Liu, J. Zhao, L. Mo, Z. Yang, J. Wang, and X. Li, “Locating sensors in the wild: pursuit of ranging quality,” in *Proceedings of Embedded Networked Sensor Systems (SenSys)*. ACM, 2010, pp. 295–308.

[4] Y. Liu and Z. Yang, *Location, Localization, and Localizability*. Springer-Verlag, 2011.

[5] H. Friis, “A note on a simple transmission formula,” *Proceedings of the IRE*, vol. 34, no. 5, pp. 254–256, 1946.

[6] Ubisense, http://www.ubisense.net.

[7] P. Coronel, S. Furrer, W. Schott, and B. Weiss, “Indoor location tracking using inertial navigation sensors and radio beacons,” in *Proceedings of Internet of Things (IOT)*. Springer-Verlag, 2008, pp. 325–340.

[8] A. Bardella, N. Bui, A. Zanella, and M. Zorzi, “An experimental study on IEEE 802.15.4 multichannel transmission to improve RSSI-based service performance,” in *Proceedings of Real-world Wireless Sensor Networks (REALWSN)*. Springer-Verlag, 2010, pp. 154–161.

[9] J. I. Huircán, C. Muñoz, H. Young, L. Von Dossow, J. Bustos, G. Vivallo, and M. Toneatti, “ZigBee-based wireless sensor network localization for cattle monitoring in grazing fields,” *Comput. Electron. Agric.*, vol. 74, no. 2, pp. 258–264, 2010.

[10] X. Wang, O. Bischoff, R. Laur, and P. Steffen, “Localization in wireless ad-hoc sensor networks using multilateration with RSSI for logistic applications,” *Procedia Chemistry*, vol. 1, no. 1, pp. 461–464, 2009.

[11] D. Puccinelli and M. Haenggi, “Multipath fading in wireless sensor networks: measurements and interpretation,” in *Proceedings of Wireless Communications and Mobile Computing (IWCMC)*. ACM, 2006, pp. 1039–1044.

[12] T. Rappaport, *Wireless Communications*. Prentice Hall, 1996.

[13] Mote Runner, http://www.zurich.ibm.com/moterunner.

[14] G. Zanca, F. Zorzi, A. Zanella, and M. Zorzi, “Experimental comparison of RSSI-based localization algorithms for indoor wireless sensor networks,” in *Proceedings of Real-world Wireless Sensor Networks (REALWSN)*. ACM, 2008, pp. 1–5.

[15] Y. Shang, W. Ruml, Y. Zhang, and M. Fromherz, “Localization from mere connectivity,” in *Proceedings of Mobile ad hoc Networking & Computing (MobiHoc)*. ACM, 2003, pp. 201–212.

[16] S. Begemann, G. van den Beld, and A. Tenner, “Daylight, artificial light and people in an office environment, overview of visual and biological responses,” *International Journal of Industrial Ergonomics*, vol. 20, pp. 231–239, 1997.

[17] N. Fergus, M. Wilson, and C. Chiancarella, “Using field measurements of desktop illuminance in European offices to investigate its dependence on outdoor conditions and its effect on occupant satisfaction, and the use of lights and blinds,” *Energy and Buildings*, vol. 38, no. 7, pp. 802–813, 2006.

A Smart City-Smart Bay Project - Establishing an Integrated Water Monitoring System for Decision Support in Dublin Bay

Fiona Regan, Timothy Sullivan, Ciprian Briciu,
Helen Cooney, Kevin Murphy
Marine and Environmental Sensing Technology Hub (MESTECH)
National Centre for Sensor Research
Dublin City University
Dublin, Ireland
fiona.regan@dcu.ie, tim.sullivan@dcu.ie
briciu.ciprian@gmail.com, helen.cooney2@mail.dcu.ie
kevin.murphy@dcu.ie

Dian Zhang, Edel O'Connor,
Noel O'Connor, Alan Smeaton
CLARITY Centre for Sensor Web Technologies
Dublin City University
Dublin, Ireland.
dian.zhang2@mail.dcu.ie, edel.oconnor@dcu.ie
noel.oconnor@dcu.ie, alan.smeaton@dcu.ie

***Abstract*—Environmental and water quality monitoring is key to measuring and understanding the chemical and biological quality of water and for taking reactive remedial action. Over the coming years, monitoring of water bodies will increase within Europe, in order to comply with the requirements of the Water Framework Directive (WFD, Council Directive 2000/60/EC), and globally owing to pressure from climate change. The establishment of high quality long-term monitoring programmes is regarded as essential if the implementation of the WFD is to be effective. However, the traditional spot/grab sampling using conventional sampling and laboratory based techniques can introduce a significant financial burden, and is unlikely to provide a reasonable estimate of the true maximum and/or mean concentration for a particular physico-chemical variable in a water body with marked temporal variability. When persistent fluctuations occur, it is likely only to be detected through continuous measurements, which have the capability of detecting sporadic peaks of concentration. The aim of this work is to demonstrate the potential for continuous monitoring data in decision support as part of a smart city project. The multi-modal data system shows potential for low-cost sensing in complex aquatic environments around the city. Continuous monitoring data from both visual and water quality sensors is collected and data from grab samples collected support the observations of trends in water quality.**

Keywords - Continuous water monitoring; estuary; marine; decision support; turbidity; salinity.

I. INTRODUCTION

The use of *in-situ* sensors capable of continuously sampling chemical and physical parameters offers the potential to reduce costs, provide more up-to-date information and a better representation of long-term trends in the fluctuations of pollutant concentrations [1] in aquatic environments. The ideal monitoring system of the near future might consist of a network of sensors deployed at key locations, capable of autonomous operation in the field for a year or more [2] [3]. Despite the increasing range and diversity of techniques currently available, continuous on-line *in-situ*, measurement systems remain largely limited by environmental factors, interferences, fouling problems, cost, power requirements, short life-time and the need for chemical reagents, as well as frequent calibrations. While the measurement and detection of environmental pollutants can be successful under laboratory conditions, continuous monitoring remains a challenge. The area of wireless sensing, and particularly, the concept of wireless networked sensors, is fast becoming one of the most dynamic and important areas of multi-disciplinary research [4] [5].

A. The requirement for monitoring

Historically, investment in the monitoring of European water bodies has been low, partly owing to the high costs associated with sample collection, and subsequent analyses in the laboratory. However, monitoring of water, globally and within Europe, will increase over the coming years, in response to the needs of the Water Framework Directive (WFD) [6] and the pressures of climate change, which will lead to resource scarcity and water quality changes. The use of relatively inexpensive *in-situ* sensors offers the potential to reduce costs considerably, making it possible to monitor an increasingly wider set of parameters in the field, as well as providing more useful, continuous monitoring capabilities to give an accurate idea of changing environmental and water quality. As mentioned previously, the accurate measurement and detection of environmental pollutants is feasible under laboratory-controlled conditions, but doing so with continuous *in-situ* monitors remains the most challenging aspect of environmental sensing. One of the advantages of wireless sensor networks is that they enable remote continuous monitoring of the environment. Data from monitoring systems can now be used for a variety of applications in addition to protection of the environment [7] [8] [9].

B. The ideal system

Although it is evident that some elements of this ideal monitoring system are in place, ongoing research and development

 ISBN: 978-1-61208-296-7

is required in several areas relating to both sensor technology and field-testing.

The ideal monitoring system of the near future might consist of a network of sensors, deployed at key locations, capable of autonomous operation in the field, for perhaps a year or more. Currently, the building blocks necessary to achieve the ideal scenario, of the measurement of multiple water quality parameters, simultaneously, in real-time are available [10]. However, as a scientific community, we need to improve the quality of some of the more sophisticated sensors for nutrients, while using the simpler devices in cleverer ways in embedded networks to make this ideal truly achievable. Data from monitoring stations can be analysed and communicated by wireless technology, for statistical processing and interpretation by expert systems, from the office. Alerts can be issued to relevant personnel - through an alarm sent to their smart phone or by e-mail - when worrying trends for any constituent of interest or breaches of Environmental Quality Standards (EQS) are detected through the evaluation of water quality parameters measured numerous times per day. These personnel can then intercept serious pollution incidents or lead the response they deem appropriate.

The aim of this work is to outline the potential for continuous water quality monitoring in decision support as part of a Smart Bay element [11] [12] to a broader connected city project in Dublin. Over the coming years, this SmartBay project will see the expansion of a multi-modal sensor and data network in Dublin bay for monitoring water quality and flooding in particular. The latter will consist of a number of sensor deployments, including visual sensing systems, modelling and additional available data sources. The data collected over the course of the SmartBay project can be utilized for other applications depending on user requirements or emerging applications, with particular emphasis on water in the city, port and coastal area.

In this paper, data from two sites in Dublin Bay will be discussed in terms of event detection and decision support opportunities. The multi-modal data system shows potential for low-cost sensing in complex aquatic environments such as estuaries. Continuous monitoring data from water quality sensors is evaluated and analysed along with data from grab samples, with the latter supporting the observations of trends from the water quality monitoring systems. The scenarios provided are a snapshot of the potential value of such a monitoring system in building a SmartBay infrastructure.

This paper introduces continuous monitoring of environmental water parameters using *in-situ* sensors in the context of a Smart City-Smart Bay project. In Section II, both of the monitoring locations used in the study are outlined. Section III describes the technology employed at the sites, how it is used and maintained as well as the methods employed to gather and analyse the environmental samples. The results obtained from the work are presented and discussed in Section IV. Finally, the conclusions garnered from this research are detailed.

II. MONITORING LOCATIONS

A sensor sonde was located at two sites in Dublin Bay (Fig. 1); Malahide Estuary (lat: $53^\circ\ 27'\ 14''$, long:$-6^\circ\ 9'$) and Poolbeg marina (lat: $53^\circ\ 20'\ 39$", long:$-6^\circ\ 13'$). Visual

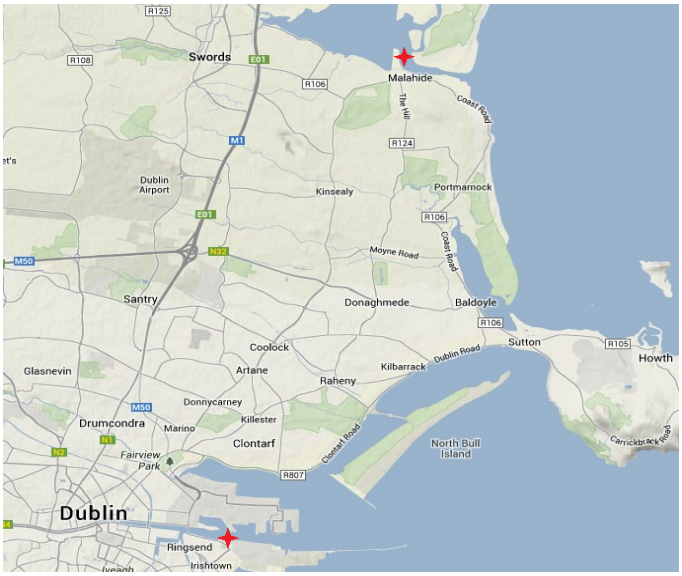


Fig. 1: Map of Dublin Bay area with the location of the working sites indicated by the red marks on the map.

sensing technology was also deployed at Poolbeg to track vessel movement in the port area. A sampling regime informed by the continuous sensor systems was instituted to collect grab samples. The two estuaries were selected for this study as they represent the variety of activities in Dublin Bay, with a wide array of stressors and user groups.

A. Malahide Estuary

As outlined by O'Boyle and Silke [13], the area around the Malahide deployment is a "typical estuarine body characterised by an elongated shallow channel demarcated at one end by a river which is the source of freshwater input and associated nutrients, and at the opposite end by the sea, which is a source of tidal seawater influence". Under the EU Habitats and Birds directives [14] [15] the estuary is both a special area of conservation (SAC) and a special protected area (SPA) and is divided in two by a viaduct built in 1844 [16]. It is a lagoon in character due to a sand spit gradually cutting it off from the sea.

The inner estuary on the left hand side of the bridge is fed with freshwater from the Broadmeadow river and the outer estuary is mainly influenced by the Irish Sea which drains almost completely at low tides. The inner estuary does not drain at low tides apart from the extreme inner estuary. As with all estuaries, the average water movement is towards the sea, but the Malahide/Broadmeadow is a mixed estuary [17] as the river flow is less dominant than the tidal flow. The main pressures affecting the estuarine catchment water quality is the presence of waste water treatment plants and sewer overflows.

B. Poolbeg Marina

Poolbeg Marina, on the lower Liffey Estuary, is a busy port environment with a diverse ecosystem that includes benthic communities [18], fish and shellfish, marine bird populations and some marine mammals. The site forms a zone of passage for salmon and sea trout migrating from the sea to spawning

 ISBN: 978-1-61208-296-7

areas upriver and for juveniles migrating from the river to the sea. The area hosts much human activity including heavy port use, marine transportation and aquaculture. The topography of the estuary is heavily modified, being walled for its whole length and undergoing regular dredging.

The site is located in the upper part of the estuary, where ship traffic is less intensive and thus easier to monitor. The water depth in the area is approximately 8 m and the width of the channel is approximately 260 m. The area acts as a buffer zone for the freshwater input and the tidal flow. Stratification is present due to denser salt water settling at the bottom with fresh water at the surface and also to seasonal heating which causes a differential event, with warm water in the surface layer isolated from the colder, bottom layer. Overlaid on these natural effects are changes attributed to anthropogenesis including input of pollutants (run-off, storm drains, sewage treatment discharges, industrial discharges, port activity and recreational boating) and the modification of flow (upstream dam releases). All of these changes effect the chemical and physical parameters at the site increasing its complexity.

III. TECHNOLOGY DEPLOYED AT BOTH SITES

As discussed earlier, a multi-parameter sonde from YSI Hydrodata UK, equipped to measure turbidity (NTU), optical dissolved oxygen (mgL^{-1}), temperature ($^{\circ}C$), conductivity (mS/cm) and depth (m), along with a telemetry system (EcoNet), was purchased and deployed. The sonde was placed at a depth of 2.5 m and data was collected with a sampling interval of 15 min. The power supply was a 12 VDC external battery and the data was recorded to an internal logger. The data was also transmitted via GPRS to a web-based server were it can be visualized in real-time or downloaded. Deployment and data collection at Malahide occurred from March until May 2012, with the Poolbeg site on-line from 1st October 2010 until the present (June 2013).

A. Maintenance Protocols

The operation of the system was checked daily by reviewing the sensor function, the battery levels and telemetry operation. Site visits were carried out for maintenance purposes where cleaning, calibration and validation measurements were performed. A ProPlus hand-held multi-parameter instrument from YSI UK was used to check temperature, dissolved oxygen and salinity and a portable turbidity meter Turb®430 IR from VWR Ireland was employed to validate turbidity measurements. These instruments were laboratory calibrated prior to site visits, which occurred every two weeks for the cold season and weekly during spring. This frequency was generally governed by the biofouling rate of sensors and, occasionally, by a malfunction of the telemetry system. Copper tape and mechanical wipers (for the optical sensors such as DO and turbidity) were used to limit the fouling rate. A protocol for the operation and maintenance of a continuous water quality monitor at sites with rapidly changing conditions was adapted from Wagner et al. [19]. This process is illustrated in Fig. 3. On arrival at the site, the following steps were performed:
1) readings were recorded from the sonde and compared to the pre-calibrated meters;
2) if a large different was observed, an insulated bucket was filled with ambient water with both the sonde and the calibrated meter placed inside and allowed to run in parallel while logging data internally;
3) the sonde and the sensors were cleansed and step 2 was repeated;
4) the sonde was removed, rinsed thoroughly and it was checked against the calibration standards and readings were recorded, in the case of the calibration criteria* being breached a re-calibration was performed.

To account of the variable nature of turbidity it's sensors pre-cleaning measurements were made in a bucket of tap water.

* calibration criteria: temperature $\pm 0.2^{\circ}C$; DO ± 0.3 mgL^{-1}; specific conductance $\pm 3\%$ of the measured value; turbidity $\pm 5\%$ of the measured value.

B. Rainfall Data

Daily rainfall data was collected from the Irish Meteorological Service from all the meteorological stations in Dublin area. Stations (Fig. 2) were selected based on the spatial distribution around the monitoring locations and the data obtained from all the stations was averaged to compute a single daily mean.

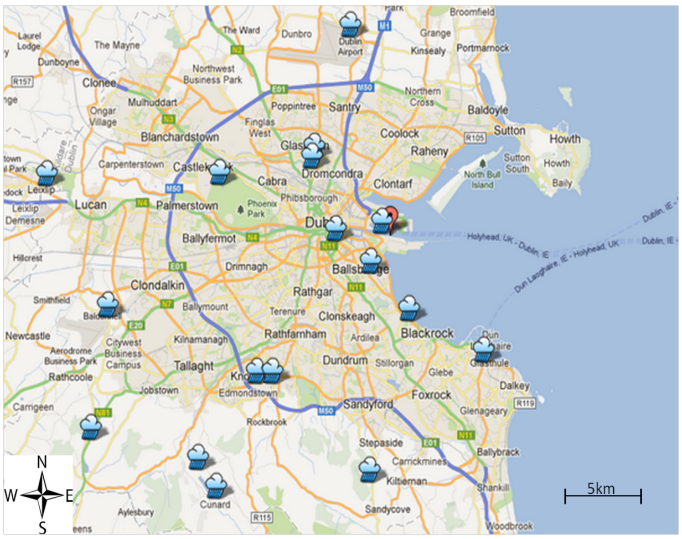


Fig. 2: The location of rainfall data stations around Dublin from which a daily mean was established.

C. Grab Samples

Water samples for phosphorus (P), Total Suspended Solids (TSS) and microbiological analysis were collected in August and September 2012. Samples were collected in the morning, before and after the arrival of the ferry at three depths: 0.5, 2.5 and 4.5 m. All the samples were collected in triplicate using a Wheaton grab sampler, and were transported to the lab under ice.

D. Microbial analysis of water samples at Poolbeg

Water samples were collected in sterile 500 mL High-Density Polypropylene (HDPP) bottles. Samples were transported to the lab under ice within 2 hrs and inoculated within 4 hrs after collection. A *Colilert-18*®/Quanti-Tray 2000® system (IDEXX Laboratories) was used for the enumeration of

 ISBN: 978-1-61208-296-7

Coliforms and *Escherichia coli* and an Enterolert ® /Quanti-Tray 2000® system (IDEXX Laboratories) was used for the enumeration of *Enterococcus* (ENT). The enumeration protocol was followed in accordance with the manufacturers instructions. Aliquots of 10 mL from the original water sample

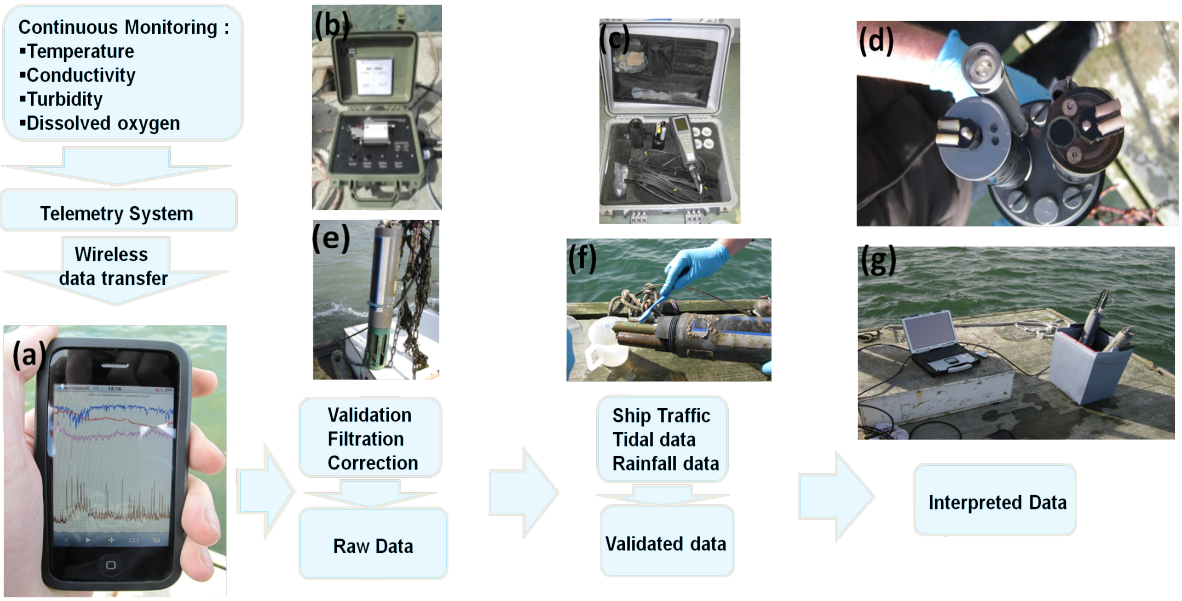


Fig. 3: Data collection, analysis and interpretation; (a) real-time data visualisation in the field using smart phones, (b) telemetry system, (c) YSI Pro-Plus hand-held sensor used for on-site validation, (d, e) YSI 6 series sonde, (f) sensor cleaning and (g) validation procedure.

were diluted 1:10 with sterile de-ionised water into 100 mL bottles. After the addition of Colilert-18 and Enterolert, samples were inoculated into Quanti-Trays and sealed. For *E. coli* and coliform enumeration, samples were incubated at 37.0°C for 18 to 20 hrs. Following incubation the Quanti-Tray wells were read for yellow colour indicating the presence of coliforms and for blue fluorescence indicating the presence of *E. coli*. For ENT enumeration samples were incubated at 41.0°C for 24 to 28 hrs after which blue fluorescent wells were counted as positive. The number of positive wells was recorded for both tests and converted to Most Probable Number (MPN) estimations using tables provided by the manufacturer. For quality control, replicates of positive controls of *E. coli* ATCC 11775 and *Enterococcus faecalis* ATCC 19433, negative controls and laboratory reagent blank were analysed for each sample batch.

E. Phosphorus analysis for Poolbeg samples

For Phosphorus (P) analysis, a protocol based on the phospho molybdenum blue method was adapted from Murphy and Riley [20] and Standard Method 4500 P-E [21] was used, as described below. Samples were collected in 500 mL HDPP bottles and transported to the lab on ice, where 250 mL of each samples was filtered through 0.45 μm cellulose nitrate filter membranes. Soluble Reactive Phosphorus (SRP) and Total Soluble Phosphorus (TSP) were determined from the filtered samples, while Total Phosphorus (TP) was determined from the raw water samples. Total Particulate Phosphorus (TPP) was determined as TP $-$ TSP $=$ TPP. For TSP and TP analysis samples were digested using acidified potassium persulfate and autoclaving for 40 min at 121°C [21]. The Method Detection Limit (MDL) was 11.0 gL^{-1} with a useful analytical range between 36.9 μgL^{-1} and 1 mgL^{-1} PO_4-P using a Shimadzu UV-1800 spectrophotometer.

F. Data management and processing

In situ sensors operating in harsh environments like those experienced at Malahide and Poolbeg are prone to fouling and drift, leading to data compromise. Before the data can be used, it has to undergo a set of quality assurance and quality control procedures to ensure that anomalies and spurious data values are removed. Wagner *et al.* [19] provide guidelines and standard procedures for correcting errors in continuous water quality data streams and Horsburgh *et al.* [22] give examples of raw data containing errors that have to be corrected. In the initial stage of this process, raw data was reviewed daily and anomalous data and out of range data were flagged [22].

Data was fully reviewed after each site visit and corrected if necessary. Out of range values were short-lived (no more than one sensor value) and were corrected manually by interpolating adjacent values. Sensor fouling occurs gradually over time leading to shifts in data after cleaning and re-calibration. Data correction for fouling drift was applied between two servicing dates as described by Wagner *et al.* [19], only when the combined absolute value for calibration and fouling error exceeded the following criteria: temperature ± 0.2°C; DO ± 0.3 mgL^{-1}; specific conductance $\pm 3\%$ of the measured value; turbidity $\pm 5\%$ of the measured value.

Sensor drift is assumed to occur at a constant rate throughout the correction, with fouling commencing as soon as the sonde is deployed in the aquatic environment. Zero correction is applied at the start of the interval, the full correction at the end of it and between these dates data is linearly interpolated.

Eq. 1 [22] was used in this case for linear drift correction,

 ISBN: 978-1-61208-296-7

$$V_c = V + (V_f - V_s)\left(\frac{T_t - T}{T_t}\right) \quad (1)$$

where V_c is the drift corrected value, V is the original measured value, V_f is the response of the sensor immediately before cleaning and validation at the end of the correction interval; V_s is the response of the sensor after cleaning and calibration; T_f is the total time interval for which the correction is applied and T is the time between the end of deployment and the time when the value is measured.

IV. RESULTS

This section provides an overview of results garnered from two estuarine sites in Dublin where continuous monitoring has been carried out over the past three years. Data from sensors and other sources, has shown that events can be identified and this can trigger the need to sample for more specialist analysis. This event driven approach to sampling is a much more reliable approach to water resource management than the current methods. These long term continuous data sets provide the basis for the development of a SmartBay dimension to the Dublin SmartCity project currently underway.

The data collected from the sites provides evidence that:
a) sensor readings (specifically turbidity and DO) can indicate the onset of a primary productivity event (algal bloom);
b) temperature and DO readings can provide an early warning of the latter event;
c) turbidity data can indicate the correct time to sample a water;
d) bacteriological and nutrient analysis can confirm the value of the turbidity readings as a decision support.

The multi-modal data system shows potential for low-cost sensing in complex aquatic environments around the city.

A. Malahide Estuary

Average surface temperatures for near shore coastal waters are influenced by freshwater run-off, as occurs at Malahide. These waters trend colder in winter and warmer in summer due to the effects of incoming freshwater which in winter is colder and in summer warmer than the corresponding sea temperatures. It can be seen that the temperature of the estuary water increases on low tide when there is freshwater run-off from the Broadmeadow River (Fig. 5(a)). Salinity can also have an effect on the solubility of gases in water (Fig. 5(b)). Salt water is more dense and becomes saturated with oxygen quicker than freshwater [21]. It would be expected for dissolved oxygen to decrease with increasing salinity, however in this situation the opposite (i.e. High DO at highest salinity) can be observed in Fig. 5(b). The reason for this is twofold; the temperature is at its lowest when DO is at its peak and the salinity differences are so small that this relationship is obscured by the tidal influences. The lowest dissolved oxygen concentration over this weekly period (and consistent with other weekly periods) occurs at the lowest salinity. This coincides with the highest temperatures, as previously established above, due to the run off from the "warmer freshwater".

It was found that chlorophyll is positively correlated to the rise and fall of the tides based on salinity (Fig. 5(c)). Photosynthesising phytoplankton off the coast of Malahide and other coastal areas move inward toward the land due to tidal movement [23]. It was observed the maximum chlorophyll concentration occurs each day on high tide when salt water and freshwater are mixed . Such daily blooms of chlorophyll could represent "Spring blooms" that occur during the months of April and May. This chlorophyll data corresponds well with the dissolved oxygen data discussed previously. Throughout the monitoring period turbidity remained relatively constant (between 0-14 NTU), however, a change was observed on 4th April 2012 (Fig. 5(d)), where it increased from the background level to 94.5 NTU. It was discovered that a linked cause of this increase could be a blue green algae (BGA) event which corresponded with other water quality parameters, as well as the onset of stormy, wet weather. The wind blowing seaward moved masses of water from the inner estuary towards the outer estuary and led to the flushing of sediments and BGA from the mudflats present in the outer estuary. Real time data analysis allowed for the verification these observations and the expected marine biological growth.

B. Event monitoring in Poolbeg - a busy estuary

In Poolbeg, on the averaged daily data set regular turbidity events occurred during the deployment period, which led to a jump in turbidity. From this study, relationships between the temporal changes in certain parameters, for instance a correlation between salinity, turbidity and DO were observed. The oscillations that are observed were not caused by water level changes. It was thought that these changes could have resulted from increased rainfall events which was confirmed by the rainfall data (Fig. 4). Other events which regularly led to increased turbidity were linked with ship traffic in Dublin Bay near the deployment site (Fig. 6). This was observed numerous times daily as vessels travelled through the port area.

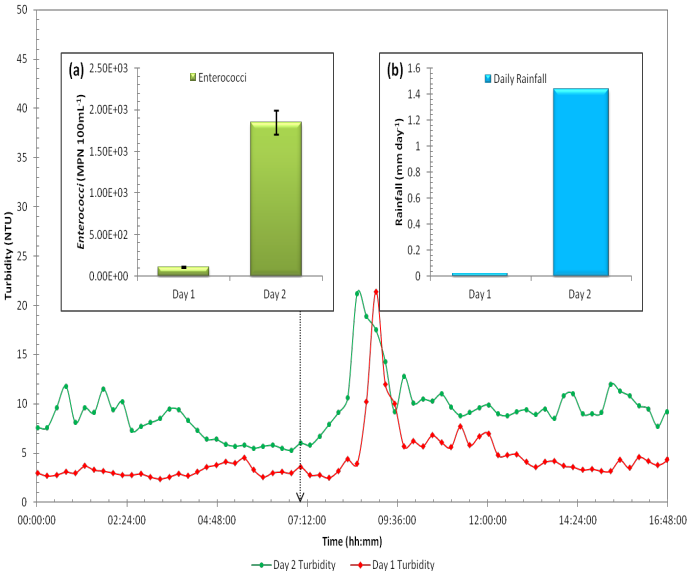
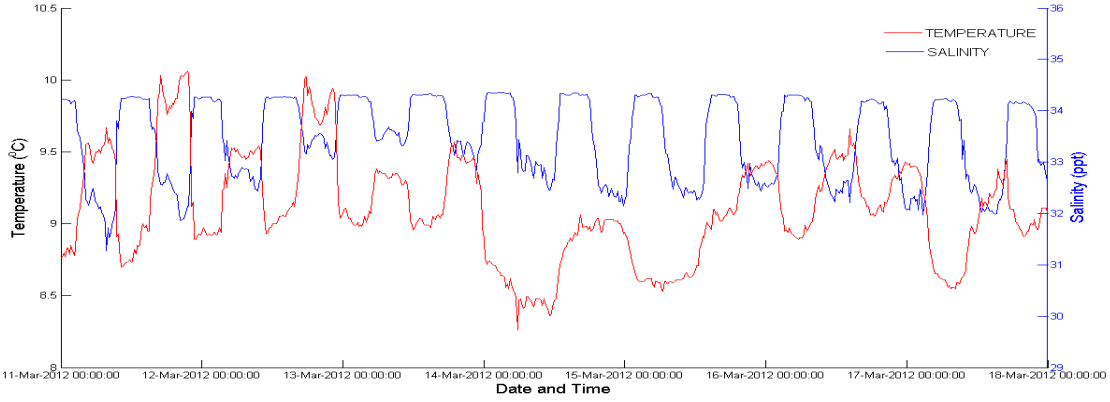
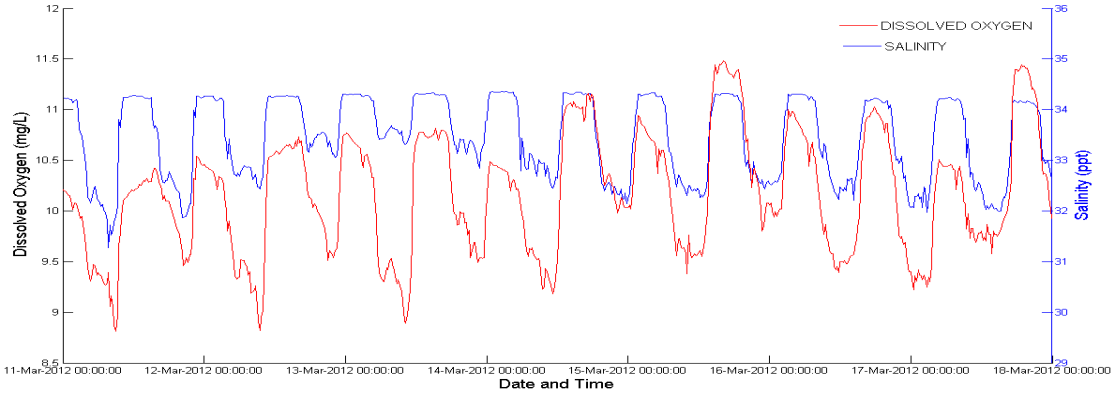


Fig. 4: Rainfall effect on background turbidity levels. Day 1 was during a relatively dry period while Day 2 was after a rainfall event. Inset (a) shows the *Enterococci* levels from the 2 days while inset (b) shows the 24 rainfall data prior to the sample collection. Discrete samples for microbiological analysis were collected on both days between 6:45 and 7:15 am.

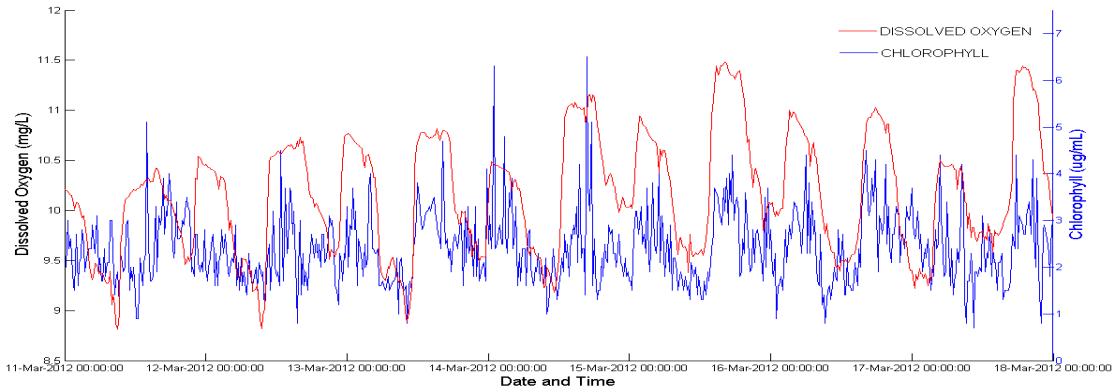
 ISBN: 978-1-61208-296-7



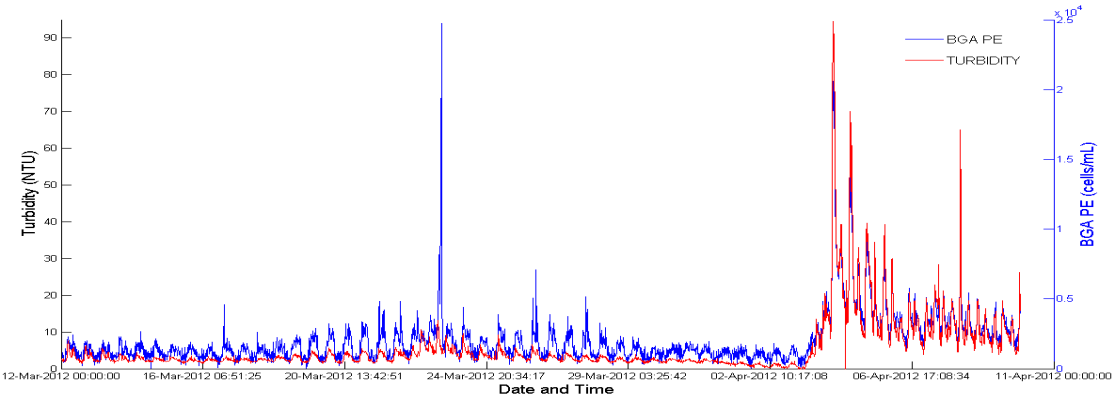
(a)



(b)



(c)



(d)

Fig. 5: Results from the Malahide estuary deployment: (a) salinity and temperature over seven days in Malahide estuary illustrating the effects of the tide and freshwater run-off, (b) dissolved oxygen and salinity overlaid indicating the daily variation, (c) overlay of dissolved oxygen and chlorophyll for a seven day period at a sensor depth of 2.5 m, illustrating the relationship between them during a growth period, (d) turbidity and BGA for a four week period showing the turbidity increase during a bloom event.

 ISBN: 978-1-61208-296-7

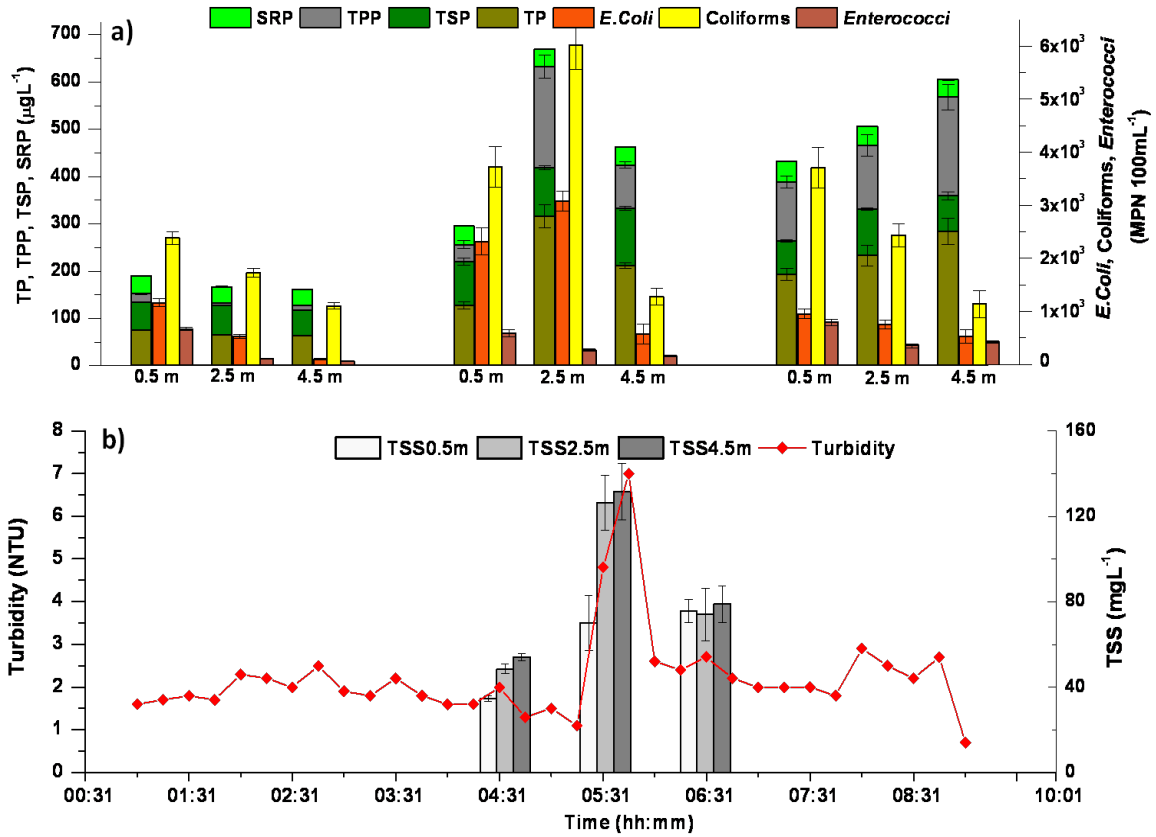


Fig. 6: Data Collected on 15th August 2012: a) Concentration of Solid Reactive Phosphorus (SRP), TPP, Total Suspended Phosphorus (TSP), TP (PO_4-P), *E. coli*, Coliforms and *Enterococci* before the arrival of the ferry (left group), immediately after the arrival (middle group) and 45 min after the arrival (right group); b) Turbidity data, TSS data before (left), at (middle) and 45 mins after (right) the arrival of the ferry. Error bars represent SD of n=3 individual samples collected at the site.

Fig. 4 shows how levels of faecal indicators in the water are elevated (Inset (a)), in association with rises in the background turbidity level. Discrete samples for microbiological analysis were collected on both days between 6:45 am and 7:15 am, with the level of *Enterococci* found to be much increased on the second day. On further investigation, it was discovered that, in the intervening day there was large rainfall, as illustrated in Inset (b) of Fig. 4.

A typical ship transit event occurred on the 15th of August 2012, as indicated by the continuous monitoring. Data grab samples were gathered at these times to identify if these events could be used to indicate ideal sampling times. In Fig. 6, the lower plot represents a turbidity increase that is linked to a vessel arrival. Samples taken before, during and after this event were analysed for microbial content (*E. coli*, Coliforms and *Enterococci*) and phosphorus (Solid Reactive Phosphorus (SRP), TPP, Total Suspended Phosphorus (TSP), TP (PO_4-P).

A concentration gradient in the water column for microbiological data was observed for the samples collected before the arrival of the first ferry, with higher bacteria densities recorded in the surface water layer (0.5 m), as seen in Fig. 6a. This was expected as the bottom layer represents a mixture of the fresh water, from the Liffey river, and the salt water. The fresh water contains higher levels of biological faecal indicators accumulated from run-off and diffuse point pollution [24]. Bedri *et al.* [25] have used a hydro-environmental model to study the impact of the Waste Water Treatment Plant (WWTP) discharges of *E. coli* on the water quality of Dublin Bay. Results from the model which accounted for wind speeds, tidal cycle, density differences within the water column and flow velocities showed that *E. coli* counts at the working site are not affected by the WWTP discharges. Another reason for the presence of the concentration gradient is the increase in osmotic pressure to which bacteria are subjected in the water column. Rich salt water environments cause the water to leave the cell and permeate through the cell wall leaving the bacteria dehydrated, eventually leading to bacterial death. For the data collected, no significant correlation was observed between the faecal indicators data and the P or TSS data, with the exception of *Enterococci*, which was moderately correlated with TSP ($R = 0.429$, $p < 0.05$).

The strongest correlation was found to be between TP and TSS ($R = 0.575$, $p < 0.05$) and TPP and TSS ($R = 0.568$, $p < 0.054$). This suggests that most of the TP enrichment in the water column is attributed to TPP associated with re-suspension of river bed material. It was identified that the major source of TP is associated with light weight sediments from the top layer of the river bed (flocculated organic material with low sheer stress).

 ISBN: 978-1-61208-296-7

V. Conclusion and Future Work

This challenging project is based on the collection of three years of continuous water quality data, supported by sample analyses for environmental parameters. The work as presented here attempts to better understand the value of continuous monitoring. It is clear from this work that simple continuous measurements of water quality parameters such as temperature, DO, turbidity and salinity can provide valuable information on the activities in estuaries around Dublin. This monitoring, supported by meteorological data, provides a valuable decision support tool which can be utilized to establish suitable water sampling times for the purpose of monitoring compliance, information about the onset of algal blooms, the frequency of pollution events and sources of pollution events among others.

Two scenarios are provided to demonstrate the potential of continuous monitoring, e.g., turbidity data as an indicator of a significant change in a system, identifying ideal times for sample collection as part of a monitoring programme and the value of salinity in identifying periods of freshwater input to the estuary, as an indicator of a potential flooding event. These data do not confirm events but rather inform a decision maker, assisting in the environmental management decision process.

The data obtained in real-time can also be valuable in monitoring the performance of the sensing system, whereby the data can identify when sensors require maintenance in a timely manner, so that the integrity of the data and monitoring system can be maintained.

The SmartBay-SmartCity project in Dublin plans to expand the sensor network in the coming year to include additional sites in the Bay. The network will include further visual sensing technologies, multi-modal data analysis and prediction tools to provide stakeholders or users with valuable information for a variety of applications in areas of tourism, water quality, security or flood management.

Acknowledgements

The authors acknowledge the funding of this research through the Beaufort Marine Research Award, carried out under the Sea Change Strategy and the Strategy for Science Technology and Innovation (2006-2013), with the support of the Marine Institute, funded under the Marine Research Sub-Programme of the National Development Plan 2007-2013. Also, QUESTOR Research Centre under grant DCU8/10/2013 and Science Foundation Ireland under grant 07/CE/I1147. The authors also wish to thank the staff at Poolbeg marina and Malahide marina for permission, assistance and access to the deployment location and marina facilities.

References

[1] R. Greenwood, J. Webster, and F. Regan, "Sustainable Water: Chemical Science Priorities. Chap 4 Water monitoring", *Royal Society of Chemistry Report*, http://www.rsc.org/images/waterreport_tcm18-108403.pdf[retrieved:August,2013], 2007.

[2] J. Grath, R. Ward, A. Scheidleder, and P. Quevauviller, "Report on EU guidance on groundwater monitoring developed under the common implementation strategy of the water framework directive", *J. Environ. Monit.*, **9**, 2007, pp. 1162-1175.

[3] A. Lawlor, F. Regan, J. Torres, B. O'Flynn, and J. Wallace, "DEPLOY: A long term deployment of a water quality sensor monitoring system", *Sensor Review*, **32**, 1, 2012, pp. 29-38.

[4] C. Albaladejo, *et al.*, "Wireless sensor networks for oceanographic monitoring: A systematic review", *Sensors*, **10**, 7, 2010, pp. 6948-6968.

[5] P. W. Rundel, E. A. Graham, M. F. Allen, J. C. Fisher, and T. C. Harmon, "Environmental sensor networks in ecological research", *New Phytol.*, **182**, 3, 2009, pp. 589-607.

[6] "The EU Water Framework Directive", *Directive 2000/60/EC*, http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2000:327:0001:0072:EN:PDF[retrieved:August,2013], 2000.

[7] S. M. Glenn, T. D. Dickey, B. Parker, and W. Boicourt, "Long-term real-time coastal ocean observation networks", *Oceanography*, **13**, 1, 2000, pp. 24-34.

[8] T. S. Moore, *et al.*, "Marine chemical technology and sensors for marine waters: potentials and limits", *Annual Review of Marine Science*, **1**, 2009, pp. 91-115.

[9] T. Voigt, *et al.*, "Sensor networking in aquatic environments-experiences and new challenges", *32nd. IEEE Conference on Local Computer Networks*, 2007, pp. 793-798.

[10] B. O'Flynn, F. Regan, A. Lawlor, J. Wallace, J. Torres, and C. O' Mathuna, " Experiences and recommendations in deploying a real-time, water quality monitoring system", *Meas. Sci. Technol.*, **21**, 2010, pp. 124004-124008.

[11] F. Regan, A. Lawlor, and A. McCarthy, "SmartCoast project- Smart-Water quality monitoring system", *STRIVE Report Series No. 30*, EPA, 2009.

[12] F. Regan, A. Lawlor, B. O'Flynn, and J. Wallace "DEPLOY: Smart demonstration of online water quality monitoring on the River Lee, Cork, Ireland", *STRIVE Report Series No. 82*, EPA, 2011.

[13] S. O'Boyle and J. Silke, "A review of phytoplankton ecology in estuarine and coastal waters around Ireland", *J. Plankton Res.*, **32**, 1, 2010, pp. 99-118.

[14] "The Habitats Directive", *Council Directive 92/43/EEC*, http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:31992L0043:EN:html[retrieved:August,2013], 1992.

[15] "The Birds Directive", *Directive 2009/147/EC* , http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2010:020:0007:0025:EN:PDF[retrieved:August,2013], 2009.

[16] R. Greene, "Around and about Malahide", *Fingal County Libraries, Co. Dublin*, 2009.

[17] D. W. Connell, "Basic Concepts of Environmental Chemistry", *Lewis Publishers, CRC Press*, 1^{st} Ed., 1997.

[18] S. Roth and J. G. Wilson, "Functional analysis by trophic guilds of macrobenthic community structure in Dublin Bay, Ireland", *J. Exp. Mar. Biol. Ecol.*, **222**, 1-2, 1998, pp. 195-217.

[19] R. J. Wagner, R. W. Boulger, C. J. Oblinger, Jr., and B. A. Smith "Guidelines and Standard Procedures for Continuous Water-Quality Monitors - Station Operation, Record Computation, and Data Reporting", *U.S. Geological Survey Techniques and Methods*, http://pubs.water.usgs.gov/tm1d3[retrieved:August,2013], 2006.

[20] J. Murphy and J. P. Riley, "A Single Solution Method for the Determination of Soluble Phosphate in Sea Water", *J. Mar. Biol. Assoc.*, **37**, 1958, pp. 9-14.

[21] American Public Health Association, American Water Works Association, and Water Environment Federation, "Standard methods for the examination of water and wastewater", *American Public Health Association, Inc.*, 18^{th} Ed., 1992 .

[22] J. S. Horsburgh, A. Spackman Jones, D. K. Stevens, D. G. Tarboton, and N. O. Mesner, "A sensor network for high frequency estimation of water quality constituent fluxes using surrogates", *Environmental Modelling & Software*, **25**, 9, 2010, pp. 1031-1044.

[23] F. Regan, *et al.*, "A demonstration of wireless sensing for long term monitoring of water quality", *The 4th IEEE International Workshop on Practical Issues in Building Sensor Network Applications (SenseApp 2009)*, 2009, pp. 819-825.

[24] B. M. Steets and P. A. Holden, "A mechanistic model of runoff-associated fecal coliform fate and transport through a coastal lagoon", *Water Research*, **37**, 3, 2003, pp. 589-608.

[25] Z. Bedri, M. Bruen, A. Dowley, and B. Masterson, "A Three-Dimensional Hydro-Environmental Model of Dublin Bay", *Environmental Modeling & Assesment*, **16**, 4, 2011, pp. 369-384.

 ISBN: 978-1-61208-296-7

Node Mobility Scheme for IP and Non-IP Wireless Personal Area Network Nodes using 6LoWPAN

Gopinath Rao Sinniah, Zeldi Suryady, Usman Sarwar, Kar Hoey Teo

IoT Lab, Wireless Communication Cluster
MIMOS Berhad
Kuala Lumpur, Malaysia
{*gopinath.rao, zeldi.suryady, usman.sarwar, kh.teo*}*@mimos.my*

***Abstract*—The rapid growth of technology has enabled more devices (IP and Non-IP, high powered and low powered) to be connected to the Internet and these devices are expected to be mobile. However, the existing IP mobility protocol is not suitable for devices that have low resources, such as Wireless Sensor Networks (WSNs). This paper proposes a mobility scheme using 6LoWPAN, targeted for any LoWPAN devices that may or may not have IP stack capability. It uses a combination of host-based and network-based mobility. In the scheme, the mobile device would be configured to store the home edge router IP address that would be given to new edge router to establish communication with the home edge router. The paper analyses the signaling messages in setting up communication once mobile node moves to a new network. The theoretical and analytical results show that the performance of the scheme is better than the existing methods.**

Keywords-Mobile IPv6; Network Mobility; Proxy based Mobile IPv6; Wireless Personal Area Network; 6LoWPAN

I. INTRODUCTION

The proliferation of devices in recent years has enabled people to expect services to be seamless even when they are moving from one network to another. Knowing this would be required, IETF has proposed MIPv6 [1] that provides IP layer handover. Since then, few more standards have emerged to cater for different solutions. Some of the mobility solutions that emerged after the introduction of MIPv6 are Fast MIPv6 [3], Network Mobility [2] and Proxy Mobility [4]. These different types of mobility solutions uses different network setup and for different use cases. It has been proven by some studies that implementations of these standards results in poor handover latency and excessive signaling [12]. This lead to a number of new schemes to improve handoff performance. It was expected that the original mobility standards could be used for wireless personal area network (WPAN) [1]; but due to its excessive signaling, new frameworks have been established.

WPAN usually consists of large number of end devices, which has very low processing, low memory and depends on battery for operation. Wireless Sensor Network (WSN) end devices, which is part of WPANs, have two components; one is the sensor used to sense some parameters and the other component is the communication interface. Once the parameter is sensed, the data is forwarded to a border router either in single hop transmission or using multihop. Single hop implementation is simple and doesn't require routing protocol as in multihop. In multihop implementation, the intermediate nodes would consume more power than end nodes because they need to route packets received from multiple end nodes. Hence, reducing the number of overhead messages would reduce some amount of power consumption.

IPv6 over low power personal area network (6LoWPAN) working group [5] was established so that these LoWPAN devices would be able to communicate with other Internet devices using IPv6 protocol. This breaks the traditional way of communication for LoWPAN that was thought incapable to connect to the Internet. Two initial proposed standards under the working group that uses adaptation layer between the MAC and Network layer have changed the way WPAN communicate with external networks. This was considered as one of the main standards to be used in the Internet of Things (IoT) technology. It is anticipated that devices in IoT would be highly mobile compared to most of the existing deployment. Thus a better mobility solution is required for both IP and non-IP devices in IoT.

Since the nodes in WPAN would have limited resources, the large code size of existing MIPv6 would be unjustifiable. Besides that, the traffic exchanges caused by MIPv6 would be very high and it would be too much for low-bandwidth wireless links. Hence, the existing methods have to be updated. There are various solutions proposed that uses either host-based or network-based mobility. This paper proposes a new way of communication for mobile WPAN devices that reduces the number of signaling in the system. The problems being tackled in this paper are:

- Existing mobility solution is not suitable to be used for LoWPAN devices hence a new approach has to be devised.
- Signaling overhead between the mobile node and the gateway is high and is not suitable for WSN nodes which depends on battery power and has low processing capability.

 ISBN: 978-1-61208-296-7

The objectives of this paper are:

- To design a method for low power personal area network (LoWPAN) devices that will have mobility support.
- To define the communication mechanism between edge routers to enable mobility for the nodes that would reduce the number of signaling messages.

The paper is organised as follows. Section II provides an overview of mobility models. Section III discusses the existing solutions specifically focusing on mobility for WPAN. Section IV provides the proposed framework and method. Section V provides the theoretical and analytical analysis while Section VI gives the implementation scenarios. Finally, conclusion and future work is presented in Section VII.

II. IPv6 Mobility

Mobility protocol for IPv6 has been divided into two broad categories, based on the role of the Mobile Node (MN) and the Edge Router. There are host-based mobility and network-based mobility.

- **Host-based mobility**: There are three components in host-based mobility; the MN, gateway (edge router) and Home Agent (HA). In this approach, the MN must have mobility stack that informs the gateway of its presence as specified in RFC6275 [1]. MN is configured to have global IPv6 address and whenever it moves to a new network, it would be configured with new address (care-of-address (CoA)) based on the prefix advertised in the network. This new address is registered at the HA so that packets destined to MN would be forwarded appropriately. HA, which resides in the home network of the MN, will intercepts packets that are destined to MN and forwards it to the new CoA of MN. MN would then directly communicates with the CN. The signaling for this host-based mobility is given in Figure 1. Since WPAN devices are resource constraint, this solution would not be appropriate as it involve lots of signaling between MN and the gateway.
- **Network-based mobility**: In this approach, the MN is not required to have mobility stack. Instead, an agent would help to inform the other components about MN's movement. This agent usually placed together with the gateway, so that whenever a new node registers, the gateway can immediately process the required information. One of the standard in this category is Proxy-based MIPv6 [4]. Few components are involved in PMIPv6; Mobile Access Gateway (MAG) that is the gateway that provide all the signaling on behalf of the MN, Local Mobility Anchor (LMA) maintains the movements of MN. When MN moves to a new network, MAG would forward MN's information to the LMA. LMA acts like the HA, which intercepts packets and forwards to the appropriate MAG using a tunnel, which was created earlier. The MAG would then sends the packet to the MN.

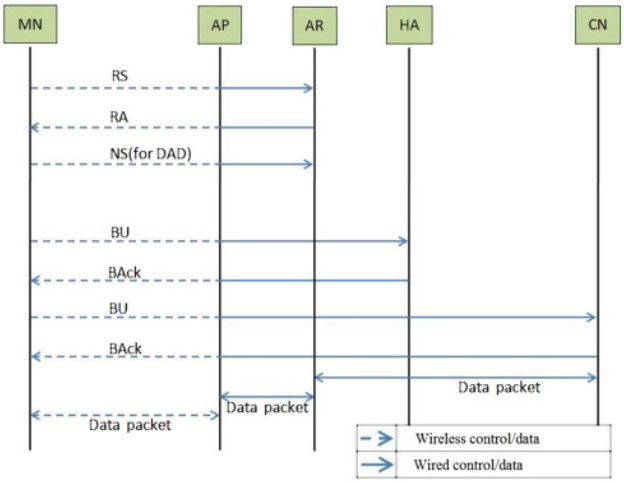


Figure 1. Signaling flow of inter-domain handover using MIPv6 [6]

Mobility of MN can also be categorized as micro mobility or Intra Personal Area Network (PAN) mobility and macro mobility or referred as Inter PAN Mobility in WPAN. In micro mobility, there is no change in the network layer address of the MN while in macro, the MN would have a new address.

Currently, there are many issues in Mobility solutions, such as high percentage of packet loss, excessive signaling, long handoff latency, high power consumption and topology issues [10]. In wireless personal area network, the issues maybe different for various types of implementations and have to be tackled on a case by case basis.

III. Existing Mobility Solutions

Due to the characteristics of WPAN nodes, the existing host-based mobility solutions are not suitable to be used. These protocols requires excessive control signals from MN to the edge router and also to the home agent. Since communication uses large amount of energy, these host-based mobility protocols are not used as it is in 6LoWPAN. To reduce the role of MN in mobility, PMIPv6 [4] was created. The mobile anchor gateway (MAG) process the control messages on behalf of the MN. When a MN moves to a new network, MAG sends the MN information to the local mobility anchor (LMA). This is achieved by using the network layer and an additional tunnel between LMA and MAG has to be established. In this way, MN does not involve in the mobility process. Each LMA would serve multiple MAGs and the same prefix can be used by MN while roaming under the same LMA. It was stated that the design of a new protocol for mobility remained an open issue

 ISBN: 978-1-61208-296-7

and further research is required to provide mobility support for resource constrained devices in WPAN [7].

Under the IETF 6LoWPAN working group, there was a proposal that provides the mobility scenarios, the main challenges and security issues [8]. There was another proposed IETF draft that provides mobility support in 6LoWPAN [9].

There are few solutions proposed for WSN based on the existing mobility stacks. In Sensor Proxy Mobile IPv6 (SPMIPv6) [12], few new components has been added to the existing PMIPv6, which are sensor LMA (SLMA) and sensor MAG (SMAG). SLMA is a powerful device that acts as topological anchor for its entire groups of SMAGs and sensor nodes, which reports to LMA. SLMA keeps binding cache entry for each sensor node in its domain and information of SMAGs table. SMAGs acts like a sink node and responsible to detect the movement of sensor nodes and handle mobility related signaling with the SLMA on behalf of the nodes. In this solution, extra devices have to be configured and added into the architecture. Another solution, based on PMIPv6 and an enhanced architecture for SPMIPv6, is cluster-based proxy mobile IPv6 for IP-WSNs [13]. In the proposed solution, the MAGs are grouped into clusters, each with a distinguished cluster Head MAG (HMAG). The new layer in the existing PMIPv6 is to reduce the load on LMA by performing intra-cluster handoff signaling and providing an optimized path for data communications.

Inter-MARIO [11], designed with a make-before-break method, is another solution that requires many static nodes in the network. In the solution, the partner node, which serves as an access point for the MN, pre-configures the future handover of the MN by sending MN's information to candidate neighbor PANs and providing neighbor PAN information to the MN. In this schemes, MN movement has to be predicted so that information pertaining to the MN can be send to the appropriate access points.

A network-based mobility that was proposed for WSN is LoWMob [15]. Similar to Inter-MARIO setup, LoWMob also uses static nodes to relay MN movement information to the gateway. Another solution that uses cluster tree that only uses the node ID to communicate with the head cluster is given in [14]. The control information interaction for both the mobility handoff and the tunnel establishment is performed at the link layer, hence it is claimed that the control information is smaller and the delay time taken to transmit this information is shorter.

IV. System Design of Node Mobility

The proposed system uses less devices to make it easier for implementation and would have lesser signaling. Design of the new proposed system is based on the combination of host-based mobility and network-based mobility. It is because both the mobile node (MN) and the Edge Router (ER) or border router has to work together to provide mobility support but the mobile node would have reduced functionality compared to system in pure host-based mobility, but unlike PMIPv6 in which MN doesn't involve in signaling. Because of this, signaling from the MN to ER would be reduced and this results in lesser power usage by the MN compared to MIPv6.

Our solution consists of MN and ERs and involves both link layer and network layer communication. Network diagram for the proposed solution is given in Figure 2.

In our solution, it is assumed that the first connected network is the home network and nodes information is stored in the Home Edge Router (HER). When a mobile node discovers its HER and associates to it, the HER will send its IP address to the MN. The HER will update its PAN table and record the details of the node. MN would also keep HER address to be used when it moves to new network.

Both HER and visited ER keep all the information of MNs in a table, which consists of the followings:

- Node ID: ID of the MN within the PAN related to the Edge Router. Since the solution is for various types of standards, the node ID has to be unique and a method for node identification and naming is being researched separately. We may combine this field and MAC field in the future as both can be used for identification and addressing.
- MAC: MAC address of the mobile node within the PAN related to the Edge Router. Since the solution is for IP and non-IP devices, some other identification system would be used such as RFID identification for RFID devices. For non-IP devices, the MAC address or unique identification would be mapped to the global IPv6 address that was given to it by the HER.
- Edge Router IP: The visited/home Edge Router's IPv4/IPv6 address.
- Home Edge Router IP: The Home Edge Router's IPv4 / IPv6 address, which the mobile node has recorded.
- Corresponded Node: A node address that communicates with the MN through HER or ER.
- MN IP: MN IP is generated by attaching the MAC address to the prefix used for LoWPAN network. The process of registering of creating this is explained 6LoWPAN Gateway given in [16].

An example of updated table at HER is given in Table I.

HER also acts as the home agent as in MIPv6, that is to capture packets from corresponding node and tunnel it to the new edge router the MN connected.

Table I
Home Edge Router Initial Table

Node ID	MAC	ER IPv6	HER IPv6	MN IP	CR
MN 1	a:a:a:a	b:b:b:b	b:b:b:b	z:z:z:z	x:x:x:x

When the MN moves to a new PAN (visited PAN) and associate to the ER, the ER of the visited PAN will record

 ISBN: 978-1-61208-296-7

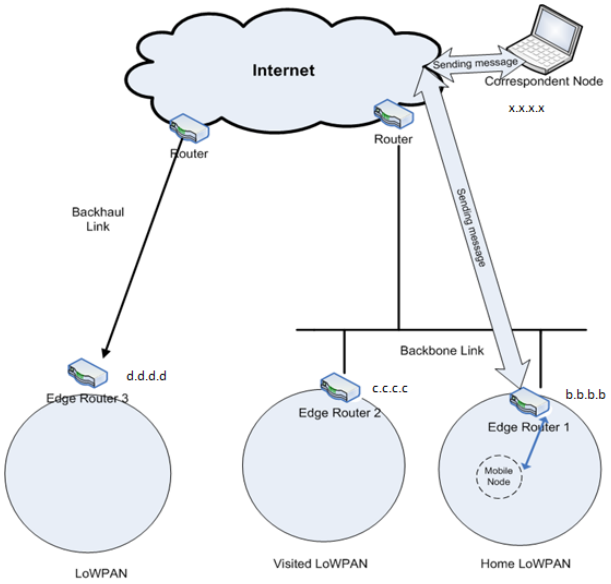


Figure 2. Overall Network consists of MN, HER, ER and CN

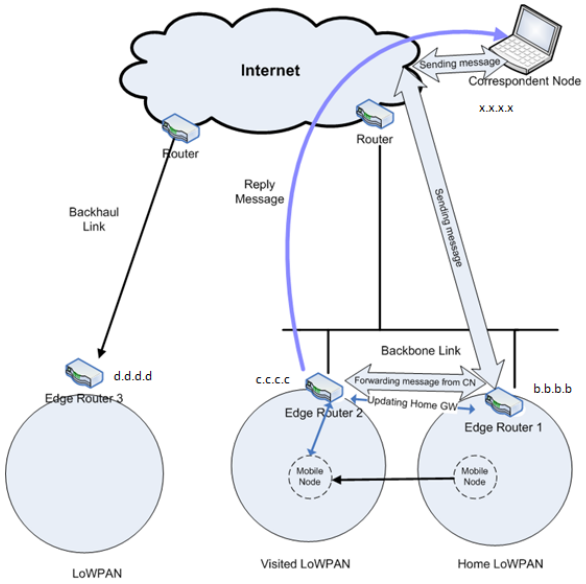


Figure 3. MN moves to new PAN

the details of the mobile node. This is executed using the methods available now, which is using layer two node association and neighbor discovery messages. Once the MN associated to the new ER, it will send the IP address of its HER to the visited ER. The ER will update its table and at the same time send an update to the HER informing that MN is now attached to the new ER. The HER will update its PAN table with the updated information. Table II shows the new information updated in new ER, while Table III shows the updated information in HER. The network diagram is represented in Figure 3 and the proposed new message flow is given in Figure 4. As shown in Figure 4, the MN attaches to new ER using the same registration as in when the MN attaches to HN. Once the link layer attachment is completed, MN would registers its HER address to the new ER. New ER, which has the HER address, would then send the location of MN and get it updated in the HER database. If there is a packet destined for MN from a CN, the HER would intercept, encapsulate the packet and then forwards it to the new ER. New ER would then decapsulate and forwards the packet to MN. This new flow of messages eliminates the binding update from AR to HA as in MIPv6.

When the MN moves to another new ER, the same process is executed. Since HER already has the ER entry in the table, HER would first inform the previous ER to remove the MN entry and then update the table with the new ER address information.

The CN will send the request messages destined for the mobile node to the HER. HER will encapsulate the request message and forward the request message to the visited PAN ER based on the details in the PAN table. The ER of the visited PAN will send all the messages from the mobile node

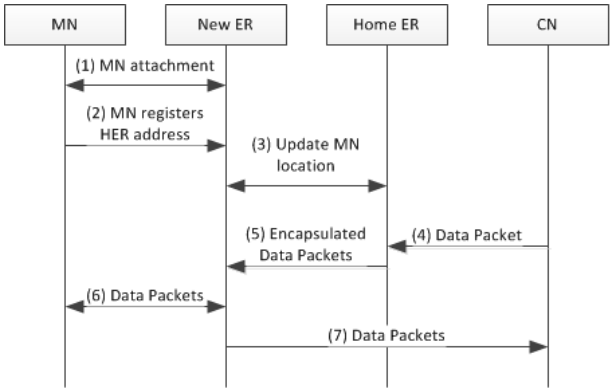


Figure 4. New Mobility Message Flow

Table II
EDGE ROUTER 2 TABLE AFTER NODE MOVED

Node ID	MAC	ER IPv6	HER IPv6	MN IP	CR
MN 1	a:a:a:a	c:c:c:c	b:b:b:b	p:p:p:p	x:x:x:x

Table III
HOME EDGE ROUTER TABLE AFTER NODE MOVED

Node ID	MAC	ER IPv6	HER IPv6	MN IP	CR
MN 1	a:a:a:a:a	c:c:c:c	b:b:b:b	p:p:p:p	x:x:x:x

to the correspondent node of the mobile node based on the PAN table.

The MN address will change every time it moves to new network, however the permanent address to outside world would be the one registered at HER. The network diagram with the address for each is given in Figure 5.

 ISBN: 978-1-61208-296-7

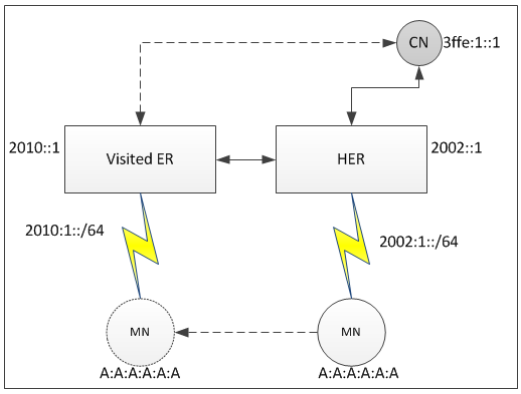


Figure 5. Network Diagram with Addressing configuration

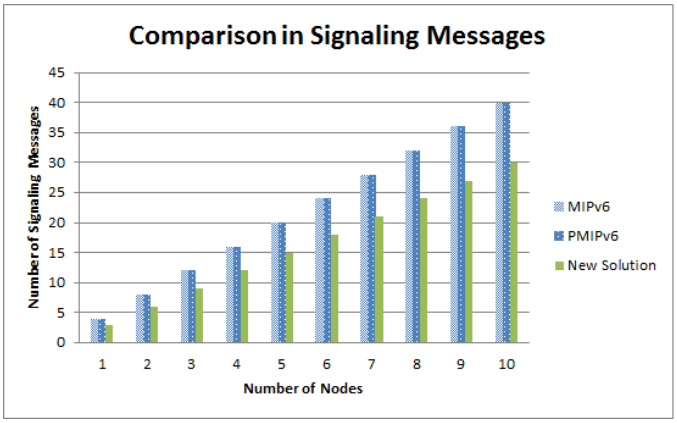


Figure 6. Comparison of Signaling Messages

V. Theoretical and Analytical Measurement

The solution that we propose is analysed in terms of signaling and scalability and compared with MIPv6 and PMIPv6.

Comparison between our solution with MIPv6:

- In MIPv6, all the signaling is from the MN starting from registration in the new network until binding to the HA as shown in Figure 1, whereas in our solution, MN only registers to the new network and all other signaling is executed by the ER on behalf of the MN. This reduces the involvement of MN in exchanging messages, which further reduces the power comsumption.

Comparison between our solution with PMIPv6:

- In PMIPv6, MN's movement is registered to the MAG, which then informs the LMA. Packets are intercepted at the LMA before being tunneled to MAG. In our solution, the ER would directly informs the HER of the MN attachment, so it can be deployed in any network and easily scalable. PMIPv6 cannot support inter-domain handover, but in our solution, new ER directly communicates with the HER in any domain it operates.

We use a simple analytical model to analyze the signaling. The following notations and assumptions are used:

t_a = the signal message from MN to ER or MAG

t_b = signal from ER to HER or HA or LMA

With the assumption that the signaling weight is the same and there is no authentication involved, the total signaling cost (sc) from the time the MN attached to new edge router until acknowledgment received from the HA or LMA or HER for MIPv6, PMIPv6 and our solution are given below. Some of the signaling cost has been combined as one unit such as for MN association.

$MIPv6_{sc}$ = signal from MN to AR (Association) + signal from AR to HA (Binding Update)+ signal from HA to AR (Binding Acknowledgement)+ signal from AR to MN (Binding Acknowledgement)

$$MIPv6_{sc} = 2t_a + 2t_b \quad (1)$$

$PMIPv6_{(sc)}$ = signal from MN to MAG (Association) + signal from MAG to LMA (Node update) + signal from LMA to MAG (Acknowledgement) + signal from LMA to MAG (Tunnel creation)

$$PMIPv6_{sc} = t_a + 3t_b \quad (2)$$

$Our solution_{sc}$ = signal from MN to ER (Association) + signal from ER to HER (Node update) + signal from HER to ER (Acknowledgement)

$$Our solution_{sc} = t_a + 2t_b \quad (3)$$

Based on the signaling messages given, the total number of signaling messages is lesser in our proposed solution compared to the MIPv6 and PMIPv6. This is shown in Figure 6 with the value of t_a and t_b as 1 each.

VI. Implementation

The system has been implemented as proof of concept for patient movement monitoring. For that, a testbed was created. We used the WSN node as the MN for patients. The nodes that are used are Sensinode WSN node which is loaded with Nanostack OS. The nodes are configured with the signaling format that was detailed earlier. Three laptops were used as the home router and edge routers. These laptops are configured to run on Ubuntu OS with Nano Router. Nano Router is an USB device that is attached to one of the available USB port in the laptop. The testbed for this implementation is given in Figure 7 .

The out patient treatment room could be the MN's home network. Edge routers can be placed in each room and whenever the patient moved out to another room, the MN

 ISBN: 978-1-61208-296-7

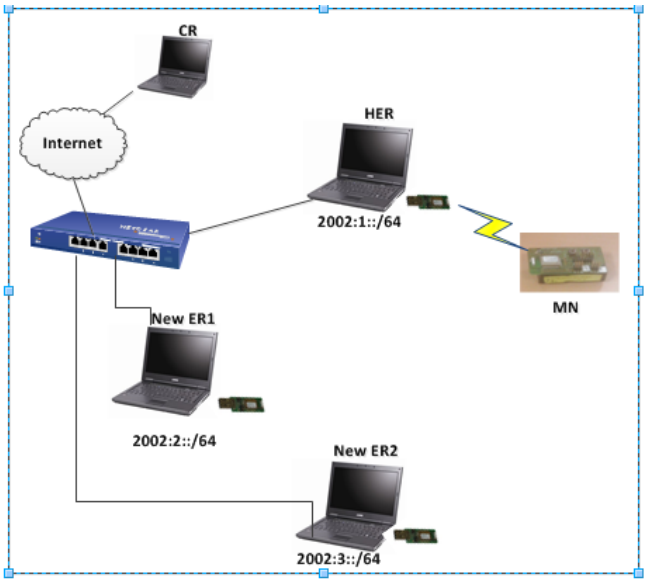


Figure 7. Testbed for validating the mobility scheme

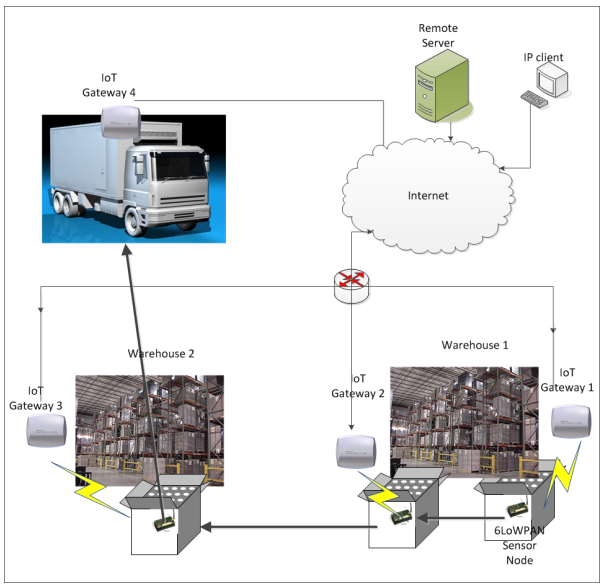


Figure 8. Implementation of Mobility solution in Logistics

would attach to the new ER, which would then inform the HER.

The mobile node successfully communicated with the correspondent node even when the MN attaches to New ER1 and New ER2. Measurements are not discussed in this paper.

Another use case that this system can be used for is for tracking goods in a warehouse until it is transported as shown in Figure 8.

VII. Conclusion and Future Work

In this paper, the new method of handling WPAN nodes mobility was introduced. The solution proposed is for both IP and non-IP devices and this would be suitable to be used in IoT scenarios, which comprises of multiple standards. The non-IP devices such as RFID, are given unique IP addresses at the gateway.

Based on the theoretical and analytical studies, the signal messages is lesser compared to MIPv6 and PMIPv6. The differences of these signaling messages is same as the total number of nodes in the network if the value for t_a and t_b are 1 each.

Future work consists of implementation of authentication between the MN and the edge router and QoS of packets. Besides that, the node ID and MAC address would be combined and used as one field with the inclusion of identification and naming method. Handover delay need to be measured as well to determine if it is better than other solutions.

Acknowledgment

This work was supported by Science Fund project (Project number: 01-03-04-SF0012) from Ministry of Science, Technology and Innovation (MOSTI), Malaysia.

References

[1] D. Johnson, C. Perkins, and J. Arkko, “Mobility Support in IPv6,” RFC6275 (Obsoletes RFC3775) (Proposed Standard), 2004.

[2] V. Devarapalli, R. Wakikawa, A. Petrescu, and P. Thubert, “Network Mobility (NEMO) Basic Support Protocol,” IETF RFC3963 (Proposed Standard), Jan 2005.

[3] R. Koodli, “Mobile IPv6 Fast Handover,” IETF RFC5268 (obsoletes RFC4068), June 2008.

[4] S. Gundavelli, K. Leung, V. Devarapalli, K. Chowdhury, and B. Patil, “Proxy Mobile IPv6,” IETF RFC 5213 (Proposed Standard),2008.

[5] IPv6 over Low Power Personal Area Network (6LoWPAN) IETF Working Group. Retrieved: March, 2013. http://datatracker.ietf.org/wg/6lowpan/

[6] K. Chaehwan, H. Hyunwoo, B. Jung-Woo, and L. Kyung-Geun, “Multicast based Proxy Mobile IPv6 for inter-domain handover. Mathematical and Computer Modelling,” Available online 10 September 2012, ISSN 0895-7177, 10.1016/j.mcm.2012.08.014.

[7] L. M.L. Oliveira, A. F. de Sousa, and J.P.C. Rodrigues, “Routing and mobility approaches in IPv6 over LoWPAN mesh networks,” International Journal of Communication Systems, Wiley, 2011.

[8] G. Mulligan, C. Williams, and D. Huo, “Mobility consideration for mobility,” Internet Engineering Task Force, Internet-Draft, draft-williams-6lowpan-mob-03.txt, July 2010, work in progress.

[9] M-K. Shin, T. Camilo, J. Silva, and D. Kaspar, “Mobility Support in 6LoWPAN,” Internet Engineering Task Force, Internet-Draft, draft-shin-6lowpan-mobility-01, November 2007.

 ISBN: 978-1-61208-296-7

[10] Teo Kar Hoey, S. Gopinath Rao, K. Reza, A. Azizol, and S. Shamala, "Mobility management schemes and mobility issues in low power Wireless Sensor Network," In Telecommunication Technologies (ISTT), 2012 International Symposium on, IEEE, 2012, pp. 65-70.

[11] H. Minkeun, K. Daeyoung, K. Seong Hoon, and H. Sungmin, "Inter-MARIO: a fast and seamless mobility protocol to support Inter-PAN handover in 6LoWPAN," In proceedings IEEE Globecomm 2010, IEEE press, New York, 2010, pp. 1-6.

[12] M.M. Islam and E.N. Huh, "Sensor proxy mobile IPv6 (SP-MIPv6) - a novel scheme for mobility supported IP-WSNs," Sensors, 11(2), 2011, pp. 1865-1887.

[13] J. Adnan, K. Shamala, Z. Zuriatu, and Nor Asilah, "A Cluster-based proxy mobile IPv6 for IP-WSNs," EURASIP Journal on Wireless Communications and Networking, 2012.

[14] W. Xiaonan, Z. Shan, and Z. Rong, "A mobility support scheme for 6LoWPAN," Computer Communications, Volume 35, Issue 3, 1 February 2012, pp. 392-404, ISSN 0140-3664, 10.1016/j.comcom.2011.11.001.

[15] B. Gargi, R. Mumammad Taqi, K. Ki-Hyung, and Y. Seung-Wha, "LoWMob: Intra-PAN Mobility Support Schemes for 6LoWPAN," Sensors, 9, pp. 5844-5877, 2009.

[16] S. Gopinath Rao, S. Zeldi, S. Usman, and A. Mazlan, "A Gateway Solution for IPv6 Wireless Sensor Network," Ultra Modern Telecommunication & Workshops, St. Petersburg, Russia, October 2009, pp. 1-6.

 ISBN: 978-1-61208-296-7

New Architecture for Efficient Data Sampling in Wireless Sensor Network Devices

Jerker Delsing, John Borg, Jonny Johansson

Luleå University of Technology
EISLAB
Luleå, Sweden
jerker.delsing@ltu.se, johan.borg@ltu.se, jonny.johansson@ltu.se

***Abstract*—When discussing powering wireless sensor network nodes, there are a few major energy consumers: communications, microcontroller and the sensor. We propose a wireless sensor network platform architecture minimizing the energy consumption of sensing. The architecture proposed herein is based on a reactive approach to sensing. A number of possible hardware approaches are evaluated and compared. This comparison indicates that analog storage between the sensing element and the sensor electronics can be a feasible method for reducing the energy consumption of the system.**

Keywords-low power WSN sensing; WSN node architecture; wireless sensor network node.

I. INTRODUCTION

One of the most common questions regarding wireless sensor networks, WSN, is what the power consumption at the sensor node must be. Much work has been done on low-powered sensor nodes and their communication abilities: see, for example, [1]–[7]. Some specific examples are schemes handling the reduction of communication [8], effective routing and multihop schemes, and the reactive partial waking up of WSN nodes [9].

Most often, the sensing element itself is disregarded from an energy budget point of view. The current state of the art for sensor interfacing is to convert the sensor data to digital form. The most common approach is using an A/D converter. Other well known approaches involves letting the sensor data influence a digital pulse train of which we easily can measure parameters like frequency, pulse width or duty cycle. None of these approaches puts the power consumption of the sensor into focus. All of these approaches are based on the assumption that all data should be transfered to some computational stage on the WSN node or any device higher up in the system architecture.

In this paper, we propose an architecture for low power interfacing of a sensing element to a WSN node. The architecture exploits the idea of detecting no or discardable changes in data. Such detection should then inhibit further processing of sensor data as early as possible. The proposed architecture is based on a reactive wake-up chain starting at the sensing element itself.

The basis for this WSN node architecture is the determination of changes in sensing element data as early as possible. In many real systems, data changes are small and are most often not of interest to the surrounding system. From an energy consumption point of view, we like to keep as much of the WSN node asleep as possible when determining whether a data sample constitutes a significant change compared with the previous sample.

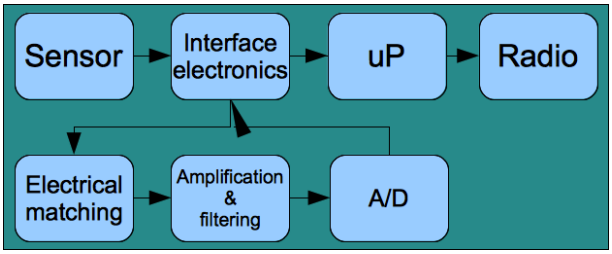


Figure 1. Traditional WSN node architecture, A1

Consider a WSN node architecture as in Figure 1. The most frequently used approach is to read the data into the μP store and compare it with the previous data. Energy is spent on sensing element sampling, signal amplification and filtering, A/D conversion and data comparison in the μP. This is the architecture you will find in most WSN and IoT (Internet of Things) nodes currently available see e.g. [10], [11].

This paper will, from an energy consumption point of view compare this generic WSN sensor interface architecture to architectures enabling early detection of discardable data changes.

A. WSN node architecture

Three different architectures targeting the early detection of useful data will be described and discussed from an energy usage perspective. As a base line, the generic architecture in Figure 1 will be used.

The first opportunity for comparison is by adding a digital memory and comparison function after the A/D conversion, as shown in Figure 2. Energy is spent on sensor sampling, signal amplification and filtering, A/D conversion and storage and comparison of the data in digital memory. Enabling the μP to sleep while performing sensor data sampling if there are no changes or small changes in sensor data will

 ISBN: 978-1-61208-296-7

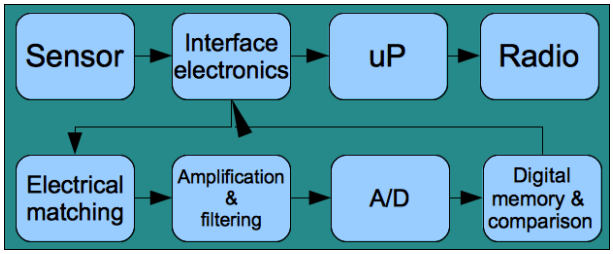


Figure 2. WSN node architecture, A2, with external digital sensor memory and comparison logic.

use energy only for the HW cost of digital memory, some configurable logic (to allow for the setting of change limits) and sending a wake-up signal to the μP.

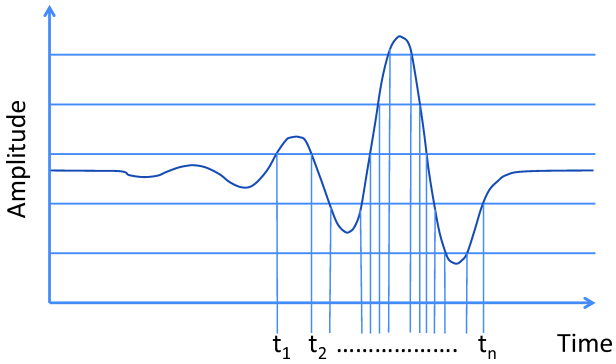


Figure 3. Illustration of level trigger sampling of a periodic signal.

Yet, one possibility is to replace the conventional A/D converter in the structures above with an Asyncronous, or Level-Crossing, ADC. As introduced in [12], [13], the level crossing ADC samples the signal when it passes a pre-determined threshold level, instead of at discrete time intervals as performed in traditional ADCs. Thus, the main output of the level-crossing ADC is the time elapsed as the signal has moved between two pre-determined levels, Figure 3. This architecture offers a couple of advantages over traditional techniques. Firstly, it allow the use of fewer quantization levels, as long as the time measurement can be performed with sufficient accuracy. Secondly, and in the context of this paper more important, the activity of the device is controlled by the behaviour of the signal that is to be sampled, allowing event-deriven data acquisition [14]. In the sensing scenario, the device will inherently idle as long as the signal stays between two threshold levels. Accordingly, no data is fed through to higher levels in the system, allowing for these to be help idle. These possibilities in the design of sensor front ends have been explored for Ultrasound measurement systems [15], as well as for other energy constrained sensor applications [16].

In terms of energy cost, the A/D conversion is dominant in the scenarios dicussed above. Thus, it is of interest to explore

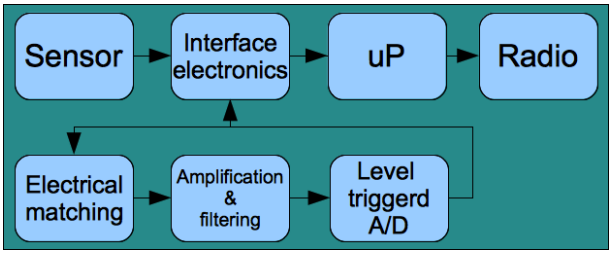


Figure 4. WSN node architecture, A3, with level triggered A/D conversion.

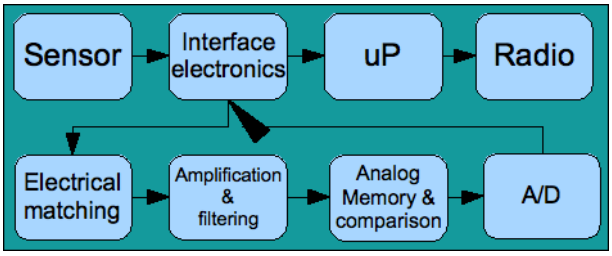


Figure 5. WSN node architecture, A4, with external analog memory and comparison logic before the A/D conversion.

the possibilities to completely remove the A/D converter. One scenario to achieve this is shown in Figure 4. Here, we introduce an analog memory with the capability of storing and comparing at least two values with associated logic and signalling to the μP. In this setup, energy is spent on sensor sampling, signal amplification, and filtering, and analog storage with its associated logic. Thus, the uP can sleep during sensning element data sampling if there are no, or small, changes in the sensed data.

Two ways to implement the analog memory are the use of Charge couple devices (CCD) and switched capacitor storage.CCDs operate by transferring stored charge packets along the surface of a semiconductor by manipulating the potentials of gate electrodes placed close to the semiconductor surface [17]. This forms essentially an analog shift register. While these devices has most commonly been used as image sensors, a number of works have over the years applied them as analog delay elements in signal processing applications.One critical advantage of using CCDs in analog delay circuits is the fact that as long as they are operated at a constant clock frequency, they are completely insensitive to matching between storage cells. In addition, the unique possibility of non-destructive readout at intermediate locations along the device have given rise to CCD based analog finite impulse response filters [18]. However, in order to move the charge packets along the surface of the semiconductor the potentials at the gates needs to change as a function of time. The charge required to affect the required change in potential is comparable to the total charge currently being transported through the device. Thus, for a CCD the power dissipation depends not only on the charge packet size required to attain

 ISBN: 978-1-61208-298-1

the required SNR, but also directly on the number of storage elements.

In contrast to CCDs, the only charge required when storing an analog sample using a switch capacitor analog memory is the charge required for the signal it self and some small charge for driving the sampling switch. The charge will stay on its capacitor without further intervention until it becomes corrupted due to leakage. It can retrieved independently on any other stored charges on other capacitors in a random-access manner if desired. Switched capacitor analog memories are however subject to the mismatch between capacitors. Both CCDs and switched capacitor analog memories present interesting alternatives to direct digitalization in systems a signal needs to be delayed before digitalization, or where it is deemed simpler do sample the signal at a high rate and later digitize the stored signal at a slower rate [19].

II. Energy analysis of architecture

The time scale on which the property being measured can be expected to change is of major importance in the choice of data acquisition architecture.

In systems designed for rapidly changing signals the time-dependent corruption of reference values stored in analog circuits (e.g., due to leakage currents) is of minor consequence and the power required for the comparators represents a significant improvement compared to continuously digitizing the signal at the rate required to capture all relevant information present in the input signal.

At time scales where storage of analog values is unfeasible, some improvements may still be attainable by using a level crossing ADC (Analog to Digital Converter) where digital registers and a DAC (Digital to Analog Converter) is used for generating the analog reference value(s).

If the signal changes sufficiently slowly the minimal static power consumption required of even the best continuous-time circuits exceed the average consuption of a regular sampling data acquisition system. Here, dedicated logic controlling the sampling process, comparing to digital thresholds and waking the microcontroller only when significant changes have occurred can yield some improvements, if the added complexity can be justified. A sampling architecture is of course also favorable in any measurement system where the sensor it self accounts for a significant fraction of the power dissipation yet can be powered down efficiently between measurements.

Below, we comment on the procedure for calculation of power consumption for the components used in the different architectures.

- Analog memory
  From the voltage noise on a capacitor at equilibrium,

$$v_n = \sqrt{k_B T / C} \tag{1}$$

  the minimal capacitor, and thus average energy per sample required for some specified SNR can be calculated. For example to reach an SNR of 60 dB with a signal amplitude of 0.5 V, a capacitance of about 30 fF is required. If unipolar charge is stored, the average energy to store samples from a sinusoidal input signal would be about 22 fJ. This compares favorably to the best ADCs in the literature (e.g., 1V 11fJ per conversion-step 10bit 10MS/s asynchronous SAR ADC in 0.18 μm CMOS consumes 11pJ per 10-bit sample). A switched capacitor memory consumes this energy once per stored value. On the other hand, in a CCD all data is shifted one step at each data acquisition. Thus, the CCD will consume an energy n times the value above per sample, where n is the number of storage positions in the CCD.
  This method gives an estimate of the theoretical minimum power required to store analog values, and is used as a lower bound.
- Time sampling ADC
  Publication of performance for this type of converter is still relatively sparse. For this paper, we use data from [15], [16].
- ADC
  ADC performance data have been collected based on [20]. The data presented in the tables below is calculated based on the therein presented Figure of merit of 100 fJ/conversion step, relevant for work published in 2011.
- Amplifier
  Amplifier data are taken from commercially available devices.
- Microcontroller
  Micro contoler data is taken from commercially available devices
- Digital logic
  Data i taken from estimates of number gates and data on commercially available gate consumption.

To analyze the energy consumption of the presented architectures, we will use four different sensing situations:

- Temperature sensing using a PT100 element as the sensing element
  - bandwidth 1 Hz
  - resolution 10 bits
- Accelerometer measurement
  - bandwidth 1 kHz
  - resolution 10 bits
- Microphone - sound recording
  - bandwidth 20 kHz
  - resolution 14 bits
- Ultrasound pulse echo measurement in liquid - piezo ceramic transducers
  - bandwidth 1 MHz - 10 MHz

 ISBN: 978-1-61208-298-1

– resolution 10 bits

This analysis is conducted using published state-of-the-art data (from our group and others) and data from commercial devices. This way, we can build an accurate picture of the total energy consumption of the proposed architectures. No circuit simulations are made here, nor have we built any complete devices.

The energy analysis is based on the following model. The total electrical power P_{tot} consumed by a WSN node can be described as:

$$P_{tot} = P_{sens} + P_{cond} + P_{A-mem} + P_{AD} + P_{D-mem} + P_{\mu P} \quad (2)$$

The total energy usage is then obtained by introducing the time needed for each operation (after which it can be turned off). To make the analysis reasonably simple, we assume that the architecture supports turning off E_{sens}, E_{cond}, once data has been stored either in analog or digital form.

For data sampling from one sensor, we assume a sensing and conditioning time t_{sens_cond}, an analog storing time t_{A-mem}, an A/D time t_{AD}, a digital memory time t_{D-mem} and a μP time $t_{\mu P}$. Thus, we obtain the total power E_{tot} used for data sampling from one sensor as:

$$E_{tot} = P_{sens} * t_{sens} + P_{cond} * t_{cond} + P_{A-mem} * t_{A-mem} + P_{AD} * t_{AD} + P_{D-mem} * t_{D-mem} + P_{\mu P} * t_{\mu P} \quad (3)$$

Provided that we have some understanding of the real values of these energies and times, we can calculate the total power consumption. In the following equation, we will do this for two sensor types, a PT-100 sensor and an ultrasonic pulse echo sensor, for each of the three architecture types A1, (see Figure 1), A2 (see Figure 2), A3 (see Figure 4), and A4 (see Figure 5).

A. Energy analysis PT-100 sensor

In this case, we assume the power consumption and time needed for a PT-100 sensor and its associated electronics, according to table I. For each of the three different architectures in Figures 1-4, we then calculate the expected span of energy consumption energy consumption based on data in Table I. The results if given in Table V. Based on the long time scales, analog memories and level triggered A/D have not been considered for this application.

B. Energy analysis accelerometer measurement

Here, we discuss the power consumption associated with the measurement of acceleration using a MEMS based 3D accelerometer. A potential usage scenario is given in Figure 6, where vibrations from passing by automotive are detected and processed. In Table II, the power consumption and related timing for the sensor and associated electronics are given. For each of the four architectures A1-A4 of Figure 1-5, we then calculate the expected span of (maximum and minimum) energy consumption based on data in Table II. The results are given in Table V.

Table I
ENERGY CONSUMPTION AND TIMING FOR A PT100 WSN SENSOR NODE.

Device	Energy consumption	Time awake [μs]
PT100	0.1-1 mW [21], [22]	10
Conditioning electronics	0.6 μW [23]	10
Analog memory	n/a	n/a
A/D (10 bit)	0.2 nW	cont
level trigged A/D	n/a	n/a
Digital memory	0.01-0.1 mW	1(storage time)
μP	1-10 mW [24]	30

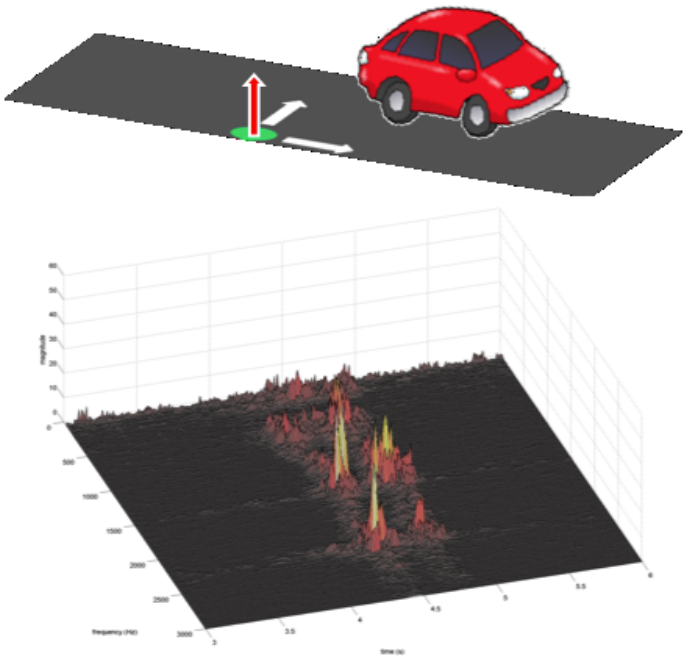


Figure 6. The application of accelerometer equipped WSN nodes for road application [25]

Table II
ENERGY CONSUMPTION AND TIMING FOR AN 3D ACCELEROMETER EQUIPPED WSN SENSOR NODE.

Device	Energy consumption	Time awake [μs]
Acelerometer	0.01 mW [26]	1 (1 μs long pulse excitation
Amplifier and filtering	1 μW [27]	10-20 (signal duration + startup time)
Analog memory	n/a	n/a
A/D (10 bit)	0.2 μW	cont
level trigged A/D	25 μW	cont
Digital memory	0.01-0.1 mW	10 (storage time)
μP	1-10 mW	30 (300 clock cycles at 10MHz)

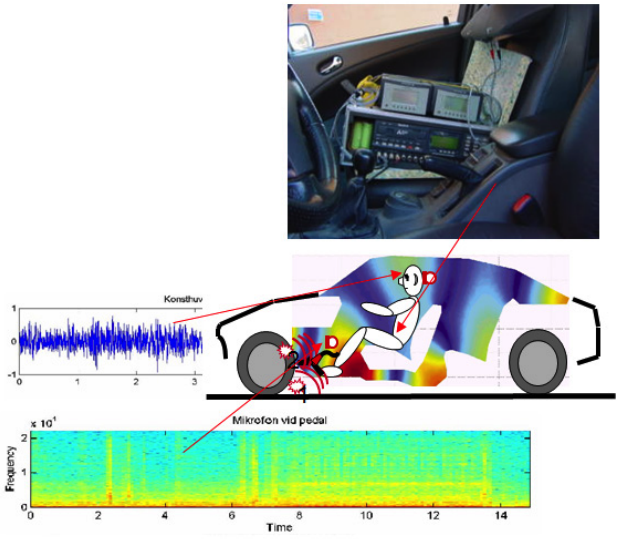


Figure 7. The application of microphone equipped WSN nodes for noise detection in automative test application

C. Energy analysis. Microphone - sound recording

Here, we discuss the power consumption and timing needed for the recording of short sounds. The WSN node is equipped with an capacitor based microphone. A potential usage scenario is given in Figure 7 In Table III, the power consumption and related timing for the sensor and associated electronics is given. For each of the four different architectures A1-A4 of Figures 1-5, we then calculate the expected span of (maximum and minimum) energy consumption based on data in Table III. The results if given in Table V.

Table III
ENERGY CONSUMPTION AND TIMING FOR AN MICROPHONE EQUIPPED WSN SENSOR NODE.

Device	Energy consumption	Time awake [μs]
Microphone	0.01 mW	1
Amplifier and filtering	0.04 mW [28]	cont
Analog memory	1 nW	cont
A/D	4 μW	cont
level trigged A/D	n/a	n/a
Digital memory	0.01-0.1 mW	10 (storage time)
μP	1-10 mW	30 (300 clock cycles at 10MHz)

D. Energy analysis ultrasound sensor 1-10 MHz

Here, we discuss the power consumption and time needed for a piezo electric transducer in an ultrasound pulse echo system used in liquid or solid media. A typical pulse echo measuring situation with typical sound signals is shown in Figure 8. In Table IV, the power consumption and related timing for the sensor and associated electronics areis given.

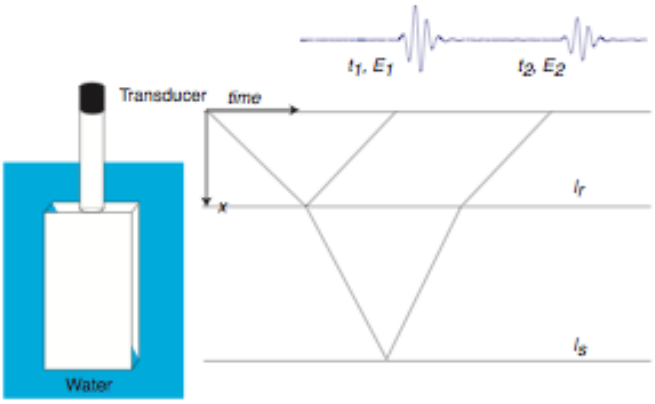


Figure 8. Ultrasound pulse echo measurement with associated acoustic signals

For each of the four architectures A1-A4 of Figure 1-5, we then calculate the expected span of (maximum and minimum) energy consumption based on data in Table IV. The results if given in Table V.

Table IV
ENERGY CONSUMPTION AND TIMING FOR AN ULTRASOUND WSN SENSOR NODE.

Device	Energy consumption	Time awake [μs]
Piezo excitation	0.01 mW [26]	1 (1 μs long pulse excitation
Amplifier and filtering	5 mW [29]	cont
Analog memory	0.5 μW	cont
A/D (10 bit)	2 mW	cont
level trigged A/D	175 mW	cont
Digital memory	0.01-0.1 mW	10 (storage time)
μP	1-10 mW	30 (300 clock cycles at 10MHz)

III. RESULTS AND DISCUSSION

Under the assumptions that we made, we have compiled power consumption data for four sensing scenarios. Data is given in table are shown in Table I til IV. The scenario operating energy consumption is computed for a 1 second period according to Equation 2 and shown in Table V.

Table V
LOW AND HIGH END ENERGY CONSUMPTION DATA IN JOULE FOR THE FOUR USAGE SCENARIOS AND THE 4 WSN NODE ARCHITECTURES.

	A1	A2	A3	A4
T_{low}	10^{-3}	$3*10^{-8}$		
T_{high}	10^{-2}	$3*10^{-7}$		
Acc_{low}	10^{-3}	$2*10^{-7}$	$3*10^{-5}$	
Acc_{high}	10^{-2}	$5*10^{-7}$	$3*10^{-5}$	
$Micro_{low}$	10^{-3}	$4*10^{-5}$		
$Micro_{high}$	10^{-2}	$4*10^{-5}$		
US_{low}	$8*10^{-3}$	$7*10^{-3}$	$2*10^{-1}$	$5*10^{-3}$
US_{high}	$2*10^{-2}$	$7*10^{-3}$	$2*10^{-1}$	$5*10^{-3}$

 ISBN: 978-1-61208-298-1

It is obvious that the μP uses a large amount of energy. Thus, any architecture that can avoid waking up the μP has clear advantages from an energy consumption point of view. This puts reactive architectures enabling the sleep fo the μP at favor.

For the discussed reactive architectures the results indicates a dividing line between sensor signals having lower frequency (~20kHz) and sensors of having higher frequency content. Another indicated dividing line is how many bits of resolution that is needed. More bits eats more power for particularly traditional A/D converters.

The following general conclusions can be made:

- For low frequency sensor signal
  - uP energy cost dominates
  - A/D + digital memory comparison is favorable
- For higher frequency sensors
  - Data conversion energy cost dominates
  - Analog memory storage and comparison has the most potential
- Reactive architecture gives in all cases an improved energy budget

Thus, it is clear that in a sensing situation with a dynamic sensor signal, such as ultrasound, avoiding A/D conversion is a promising approach. If analog memory and comparison techniques can be developed similar to what we have seen for A/D converters, analog storage architecture will be a strong contender for future WSN designs.

IV. Conclusion and further work

The reactive architecture proposed here for minimal energy consumption of sensing on WSN platforms is promising. Based on an analysis of current state-of-the-art sensor interface electronics, an approach using analog storage provides interesting data when compared with more mature technology such as ADC and ADC combined with memory logic.

Future work will reveal whether if a reactive architecture based on an analog storage approach will show improvements in energy consumption similar to those of advanced ADC. If such improvements are shown, the analog storage approach has clear merit for use in future WSN node designs.

Acknowledgment

The authors would like to thank the ESIS project [30] supported by EU structural funds the Artemis Arrowhead project [31] for financial support.

References

[1] W. Heinzelman, A. Chandrakasan, and H. Balakrishnan, "Energy-efficient communication protocol for wireless sensor networks," in *Proc. Hawaii Intl. Conf. System Sciences*, Hawaii, Jan 4-7 2000, pp. 3005–3014.

[2] J. Eliasson, M. Lundberg, and P. Lindgren, "Time synchronous bluetooth sensor network," in *Proc- IEEE Concumer Communication and Networking Conference, CCNC*, 2006.

[3] M. Lundberg, J. Eliasson, L. Svensson, and P. Lindgren, "Context aware power optimization of wireless embedded internet system," in *Proceedings IEEE IMTC*, 2004.

[4] J. Eliasson, P. Lindgren, J. Delsing, S. J. Thompson, and Y.-B. Chen, "A power management architecture for wireless sensor nodes," in *Proc IEEE Wireless Communication and Networking Conference, WCNC*, 2007.

[5] V. Loscri, G. Morabito, and S. Marano, "A two level hierarchy for low energy adaptive clustering hierarchy (TL-LEACH)," in *Proc. 62nd IEEE Vehicular Technology Confere)nce (VTC-Fall*, Dallas, 25-28 September 2008, pp. 1809–1813.

[6] Y. Xu, J. Heidman, and D. Estrin, "Geography-informed energy conservation for ad hoc routing," in *Proc. Mobicom*, 2001, pp. 70–84.

[7] J. Delsing and P. Lindgren, "Sensor communication technology towards ambient intelligence, a review," *Meas. Sci. Technol.*, vol. 16, pp. 37–46, 2005.

[8] J. Lu, F. Valois, M. Dohler, and M.-Y. Wu, "Optimized data aggregation in wsns using adaptive arma," in *Proceeedings Sensorcomm 2010*, 2010, pp. 115–120.

[9] S. G. Hong, N. S. Kim, C. S. Pyo, and W. W. Kim, "Hybrid sensor module and data processing using low-power wakeup in wsn," in *Proceeedings Sensorcomm 2010*, 2010, pp. 191–195.

[10] EISTEC. (2013) Eistec. [Online]. Available: http://www.eistec.se

[11] Memsic. (2011, Dec) Wireless modules. [Online]. Available: http://www.memsic.com/wireless-sensor-networks/

[12] N. Sayiner, H. Sorensen, and T. Viswanathan, "A level-crossing sampling scheme for a/d conversion," *Circuits and Systems II: Analog and Digital Signal Processing, IEEE Transactions on*, vol. 43, no. 4, pp. 335 –339, apr 1996.

[13] E. Allier, G. Sicard, L. Fesquet, and M. Renaudin, "A new class of asynchronous a/d converters based on time quantization," in *Asynchronous Circuits and Systems, 2003. Proceedings. Ninth International Symposium on*, May 2003, pp. 196 – 205.

[14] Y. Tsividis, "Event-driven data acquisition and digital signal processing a tutorial," *Circuits and Systems II: Express Briefs, IEEE Transactions on*, vol. 57, no. 8, pp. 577 –581, aug. 2010.

[15] K. Kozmin, J. Johansson, and J. Delsing, "Level-crossing adc performance evaluation toward ultrasound application," *Circuits and Systems I: Regular Papers, IEEE Transactions on*, vol. 56, no. 8, pp. 1708 –1719, aug. 2009.

[16] M. Trakimas and S. Sonkusale, "An adaptive resolution asynchronous adc architecture for data compression in energy constrained sensing applications," *Circuits and Systems I: Regular Papers, IEEE Transactions on*, vol. 58, no. 5, pp. 921 –934, may 2011.

 ISBN: 978-1-61208-298-1

[17] R. Baertsch, W. Engeler, I. Goldberg, H.S., C. Puckette, and J. Tiemann, "The design and operation of practical charge-transfer transversal filters," *Solid-State Circuits, IEEE Journal of*, vol. 11, no. 1, pp. 65 – 74, feb 1976.

[18] D. D. Buss, D. R. Collins, W. H. Bailey, and C. R. Reeves, "Transversal filtering using charge-transfer devices," *Solid-State Circuits, IEEE Journal of*, vol. 8, no. 2, pp. 138–146, 1973.

[19] G. Haller and B. Wooley, "A 700-mhz switched-capacitor analog waveform sampling circuit," *Solid-State Circuits, IEEE Journal of*, vol. 29, no. 4, pp. 500 –508, apr 1994.

[20] B. Murmann. Adc performance survey 1997-2013. [Online]. Available: http://www.stanford.edu/~murmann/adcsurvey.html

[21] Pentronic. Resistance thermometer theory. [Online]. Available: http://www.pentronic.com/Theory/Pt100sensor/tabid/188/language/en-GB/Default.aspx

[22] ——. The effect of 2,3 or 4 wire connection using pt100/rtds. [Online]. Available: http://www.pentronic.com/Portals/0/PDF/En/Useful%20links%20pdf/The_effect_of_2_3_4_wires_on_Pt100_060210.pdf

[23] Lpv521, nanopower op. [Online]. Available: http://www.ti.com/product/lpv521

[24] Data sheet m16c. [Online]. Available: www.renesas.com

[25] W. Birk, J. Eliasson, P. Lindgren, E. Osipov, and L. Riliskis, *Road surface networks technology enablers for enhanced ITS*. IEEE, 2010, pp. 152 – 159.

[26] J. Johansson and J. Delsing, "Energy and pulse control possibilities using ultra-tight integration of electronics and piezoelectric ceramics," in *Proc. UFFC*, vol. 3, 2004, pp. 206–2210.

[27] (2009) Microchip mcp6141. [Online]. Available: http://ww1.microchip.com/downloads/en/DeviceDoc/21668d.pdf

[28] Maxim max9914. [Online]. Available: http://datasheets.maximintegrated.com/en/ds/MAX9914-MAX9917.pdf

[29] Texas instruments opa683. [Online]. Available: http://www.ti.com/product/opa683

[30] [Online]. Available: www.esis.se

[31] [Online]. Available: www.arrohead.eu

 ISBN: 978-1-61208-298-1

Energy Evaluations for Wireless IPv6 Sensor Nodes

Cedric Chauvenet
Watteco Inc.
La garde, France
C.Chauvenet@watteco.com

Bernard Tourancheau
LIG, UMR 5217
Grenoble University, France
Bernard.Tourancheau@imag.fr

Denis Genon Catalot
LCIS, EA 3747
Grenoble University, France
dgc@iut-valence.fr

***Abstract*—Typical Wireless Sensor Networks (WSN) include some energy autonomous nodes and main powered router nodes. To meet a reasonable lifetime, i.e., years, the autonomous devices have to encompass strong energy constraints. Our target network architecture consists in such wireless autonomous nodes connected to a backbone based on Power Line Communication (PLC) nodes. Our given building automation application scenario runs for a 10 minutes periodic probes sampling, which reports a data frame over the Internet using IPv6. For this purpose, our WSN nodes leverage on open RPL/ContikiOS softwares and typical off the shelf electronic components. In order to optimize the node energy consumption we propose a systemic analysis that includes all the software layers and hardware components. The energy distribution among the different components and the critical software parameters are weighted against the global energy consumption. Thanks to measurements and technical data, we propose a simple model that allows to identify pitfalls and determine optimized solutions. Following our established guidelines, we believe our future WSN monitoring platform will achieve more than 10 years lifetime with the target scenario.**

***Keywords*—Wireless Sensor Networks (WSN), Energy Optimization, Energy Measurement, RPL, IPv6, PLC.**

I. INTRODUCTION

WSN are called to new challenges given their energy constraints. To meet their power, price, size and deployment requirements, typical platforms fostered on a design based on a low power micro controller unit (MCU) and a low power RF transceiver. Adding a battery and probes to this communicating platform leads to a cheap autonomous WSN node. WSN's lifetime should be long enough, i.e., years, to provide a reasonable return on investment. Any costly maintenance like battery replacement annihilates the interest in WSN deployments. WSN devices will be sleeping most of the time and periodically wake up to perform their operations. This is controlled by their embedded operating software stack. Though carefully designed, these softwares should be carefully tuned with respect to the application requirements. In this work, we review the energy optimization process of a node that is compliant with the latest IETF recommendations for WSN. The target application is home and building automation.

Section II states the art of the domain and Section III presents the context of our work. In Sections IV and V the power consumption study of several MCU and transceivers are presented. Section VI shows the impact of the routing parameters. In Section VII, the global RF node energy spending is optimized. Sections VIII and IX present software optimizations for the MCU and the transceiver. Section X shows the probes consumptions that will be added to the node.

II. RELATED WORK

Energy consumption is the most important criterion for the development of autonomous sensor network nodes. Especially when they target several decades or perpetually powered systems. As mentioned in [1], battery replacement is not an option for networks with thousands of physically embedded nodes and the paper list various and commonly used techniques to save power such as power-aware computing, energy-aware software or power management of radios.

Energy has been considered since the very beginning of WSN developments, like with the WeC node designed in 1998 [2] that concretizes the idea of communicating sensors. Then, some similar platforms have been designed, with more powerful MCU and the same radio chip until 2001, where a new generation of nodes emerged with the Mica [3], designed in Berkeley in 2002. This node was carefully designed to serve as a general purpose platform for WSN research. Work in [3] highlights the need for node sleeping most of the time with periodic wake up, and shows that a total improvement factor of 11 can be reached with this technique. From this date on, all the sensor network node designs used this duty cycling techniques to save power. Leveraging on experimental feedback, Mica2 corrected many shortcomings: the boost converter was discarded, new components were chosen. But, this new design leads to higher power consumption. Continuing the evolution, MicaZ replaced the CC1000 radio with a CC2420, an IEEE 802.15.4 2.4GHz compatible radio with 250 kbps throughput capability, and evolved O-QPSK modulation. This transceiver embedded a part of the 802.15.4 standard, limiting the software processing needed to send and receive packets by DSSS signal treatment, thus saving energy.

A major step has been reached with the design of the Telos platform [4] released in 2004. It enables experimentations with minimal power consumption, a maximum of 10 kB RAM memory for managing increasing stack fonctionnalities and increased software and hardware robustness. Telos is based on a MSP430 MCU with a CC2420 radio. The MSP430 offered significantly lower consumption, reducing the total active power down to 41 mW, with the same transmitting power consumption as MicaZ. According to [4], the power consumption of the MSP430 is 20 times lower than the MicaZ in sleep mode, and 4 times lower in active mode.

However, even if hardware greatly improved over time, software requirements changed, especially for the communication stack. The IETF proposed the adoption of IPv6 over WSN with an adaptation layer called 6LoWPAN, RFC 6282 [5] and a routing proposal named RPL, RFC 6550 [6]. 6LoWPAN offers header compression to save bytes and allows frame fragmentation to resolve MTU issues. The Routing Protocol for

 ISBN: 978-1-61208-296-7

Low power and Lossy Networks, RPL, is a proactive distance vector protocol that creates Destination Oriented Directed Acyclic Graphs (DODAG). Several control messages protocols are defined in the RPL's RFC. Namely, DAG information object, DIO, and information solicitation, DIS, propagates and requests DODAG informations. Destination advertisement object, DAO, allows routers to update their tables. Also, a Neighbor Discovery mechanism, RFC 4861 [7], ensures neighbors' reachability within the topology by neighbor solicitation and advertisement, NS/NA, exchanges. Notice that node energy can be used in the metric governing this routing topology construction.

This software stack complexity impacts the key parameters governing power requirement. Work in [8] evaluates the performance of RPL and 6LoWPAN using the TinyOS stack. They showed that RPL parameters' values and the number of downwards stored routes affect the power consumption. Also they proposed an optimization for fragmented packets where only the header part is decompressed for routing decisions. [9] proposed a power consumption model that deeply describes the wireless communication. Hence, this work shows that the number of hops should be reduced to the minimum in order to saved power. To our knowledge, the impact of the RPL Contiki stack on energy consumption was not studied.

We address the power consumption issues by conducting real hardware measurements and analysis of the software impact on consumption. As suggested in [1], we split the system into several parts and conduct an analysis for each of them in the first part of the paper. However, the system should also be considered as a whole in order to insure a good integration between hardware and software components. This is the aim of the second part of our paper where we optimize the software behavior against energy.

III. Platform settings

This paper relates our research in order to define an operational and efficient WSN platform using Power Line Communication (PLC) [10], and Radio Frequency (RF) communication transceivers. These architectures gathers autonomous battery powered RF nodes, main powered PLC or RF nodes and PLC-RF routers. The PLC-RF routers serve as a backbone for RF nodes. Such an architecture restricts the routing capabilities to mains powered devices. Moreover, battery operated devices do not waste energy. The gain of these energy efficient hybrid architectures is described in [11]. At the network level, our nodes are seamlessly interconnected using the Internet IPv6 protocols [12].

In order to operate the nodes for this study, we used ContikiOS and tools [13], which provide an open source micro operating system for constrained devices.

The RF transceivers follows the IEEE 802.15.4 standard [14], which was designed for low power, short range and low throughput networks. Initially released in 2003, this standard offers a maximum throughput of 250 kbps. New revisions of this standard added other modulations scheme and a sub-GHz RF band for less attenuation transmission with lower throughput, typically 20k to 50Kb/s, to achieve a longer range and more robust communications.

The target MAC layer relies on the CSMA/CA mechanisms required by the 802.15.4 standard. Also, as in most of WSN protocol stacks, a radio duty cycling (RDC) protocol is added in order to save a significant amount of power. We use an adapted version of contikiMAC RDC [15] without the preamble sampling protocol. This protocol induces periodic radio wake up to sense the radio activity at a fairly high frequency, 4 Hz by default. This protocol may sends packet copy until the recipient acknowledge it. Notice that if the traffic required by the application is lower than 4 Hz, the periodic wake up will induces non necessary overhead. Our target temperature monitoring application typically requires one packet per node every 10 minutes send to the sink.

On top of an IPv6 communication stack, our target application uses the IPSO application framework [16] profiles designed primarily for smart energy applications. These profiles relies on the Constrained Application Protocol, CoAP [17], that enables sensing nodes to be interrogated through RESTfull primitives.

IV. MCU Power Consumption

In a node, the MCU wakes up periodically to check if an event is detected by the probe or the RF transceiver, and executes the corresponding software tasks. Otherwise, it stays in sleep mode. Table I compares the current drawn for 3 MCUs in different states. The results shows the great gap between active and low power modes. For instance, the MSP430f5438A running at 20 MHz can, in theory, achieve a 5 years lifetime over an 1Ah battery only if it stays in sleep mode more than 99,6% of the time. The MCU frequency impacts its

TABLE I. Micro controllers drawn current under 3V at different modes and frequencies. Experimental measurements are in bold.

Mode	MHz	MSP430 f1611 (mA)	MSP430 f5438A (mA)	SIM3C1xx (mA)
Active	4	2 [18]	**1.51**	
Active	8	4 [18]	**1.84**	
Active	16		**5.21**	
Active	20		6.37 [19]	7.8 [20]
Active	80			22 [20]
Low P	0.032	0.002 [18]	0.0021 [19]	0.0008 [20]
Sleep		0.002 [18]	**0.0012**	0.000145 [20]

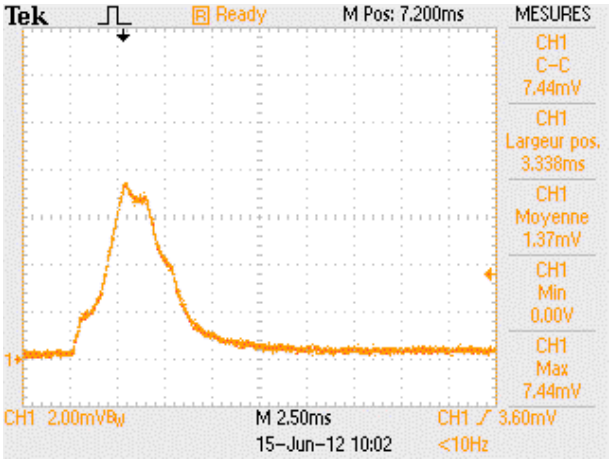
energy consumption. In order to quantify this, we measured the voltage on a 10,1 Ω load added between a regulated DC power source of 3 V and the target architecture. In the experiments, this voltage reflects the current drawn by the MCU plus the ATRF212 transceiver in its standby mode, which is 0.2 μA in [21]. We ran the complete software stack over different frequencies without any RF communication nor connected probe. Figure 1 shows the corresponding oscilloscope traces, averaged over 128 wake up samples to include the variety of wake up profiles. We calculate the energy, E by integrating the power over time during the period of visualization following (1).

$$E(J) = \frac{3}{10.1} \times \int_0^{t_{visualization}} V_{average} \qquad (1)$$

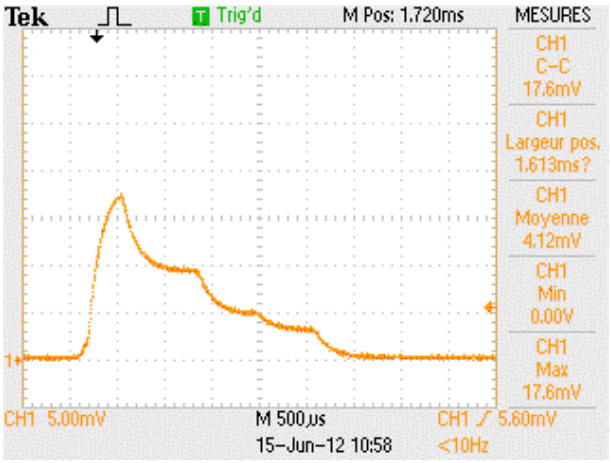
Table II summarized the corresponding computed energy.

As expected, the maximum current drawn increases with the MSP frequency while the rise up and drop time decrease,

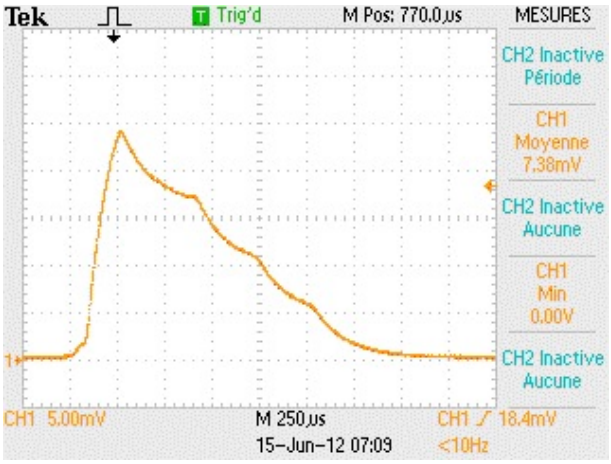
 ISBN: 978-1-61208-296-7



3.9 MHz; $H_{scale} = 2500\mu$s; $V_{scale} = 2$mV



8 MHz; $H_{scale} = 500\mu$s; $V_{scale} = 5$mV



16 MHz; $H_{scale} = 250\mu$s; $V_{scale} = 5$mV

Fig. 1. Current profile of MSP430 average wake up at different operation frequency.

TABLE II. MEASURED ENERGY PARAMETERS OF MSP430 WAKE UP PERIOD AT DIFFERENT FREQUENCIES.

Frequency of the MSP (MHz)	3.9	8	16
Rise Up Time (ms)	2.5	0.5	0.25
Maximum current (mA)	0.74	1.74	2.38
Drop Time (ms)	7.5	2.5	1.5
Time of Observation (ms)	25	5	2.5
Average Voltage (V)	1.37	4.12	7.83
Energy (μJ)	**10.2**	**6.12**	**5.48**

although none of these happens in a linear fashion. But we notice a higher gap between 3.9 and 8 MHz than between 8 and 16 MHz for these three parameters. Regarding the maximum current, there is a 1 mA gap between 3.9 and 8 MHz, and a 0.64 mA gap only between 8 MHz and 16 MHz. Notice that our measurements differ from the data sheet maximum current consumption. This can be explained for 2 reasons. Firstly, they are averages over 128 samples. Secondly, the wake up duration is not long enough to reach the maximum current consumption. The rise up and drop time are divided by a factor of 1.7 and 2 when going from 8 to 16 MHz while there are respective factors of 5 and 3 when moving from 4 to 8 MHz. The erratic performances of the MCU, while down clocked at 4 MHz, may explain this non linear behavior. Taking into account these results, the SIMC3C1 seems promising for our future platform.

V. TRANSCEIVER POWER CONSUMPTION

With our architecture, the RF nodes do not route messages. An important parameter is thus the transmission range to ensure that every RF node will reach a main powered router. The studied transceivers follow the IEEE 802.15.4 standard, which proposes several frequency bands. We focused on the 868 MHz and 2.4 Ghz ones. In order to quantify the theoretical advantage of each transceiver for our application requirements, we computed the power needed to communicate within a range d. The Friis formula states this received power as a function of the distance and power gain.

$$P_{received} = 22dB + 20 * log(d/\lambda) \quad (2)$$

Table III presents the results. The ATRF230 has a low instantaneous power consumption and a high throughput, confirmed by its "Energy at max-range" number. However, its coverage range is small compare to the others. The ATRF212 and the CC1120 transceivers have fairly similar "Energy at max-range" consumptions, because the higher power consumption of the latter in balanced by its higher throughput capability. Though, the 2.4GHz band offers a shorter range than the two others and the maximum power of transmission authorized in this band is lower. As a result, for a RF node in our architecture, a router may be out of range in 2.4 GHz while connected in 868 MHz.

Table III also shows that the computed maximum reachable distance over 868 MHz is 27 times greater than over 2.4 GHz using the transceivers and regulation considered. The energy cost per bit in much lower is 2.4 GHz for distances shorter than the 1.5km range. Also, for the same transmitted data, the channel occupation is longer in 868 MHz and the interference range is larger. This may become an issue in dense environments. Beside energy considerations, the 868MHz band may increase the path diversity while being able to connect to

 ISBN: 978-1-61208-296-7

TABLE III. CHARACTERISTICS OF RADIO TRANSCEIVERS AND CORRESPONDING ENERGY FOR DIFFERENT TRANSCEIVERS.

Module	Rate (kbps)	Freq (MHz)	Sensib (dBm)	Current Tx (mA)	Current Rx (mA)	Max Power (dBm)	Max Range (km)	Power Recv (dBm)	Energy at Max Range (μJ/bit)	Energy at 1.5 km (μJ/bit(dBm))
ATRF212	20	868	-110	25	9.2	10	25	-109.19	5.13	2.96 (-15)
CC1120	50	868	-110	45	22	14	40	-109.27	4.02	2.52 (-15)
SI4464	40	868	-110	37	13	14	40	-109.27	3.75	2.33 (10)
SI4461	40	868	-110	33	13	14	40	-109.27	3.45	2.33(10)
ATRF230	250	2400	-101	16.5	15.5	3	1.5	-100.58	0.38	0.38(3)

several routers as this allows for multi and backup routes. With our architecture and target application, we decided to use the 868 MHz band and thus the SI4461 transceiver seems a good candidate. Of course, additional parameters could be added to this model and in particular, the gain and type of antennas that could greatly impact the transmission range.

VI. ROUTING CONTROL SOFTWARE IMPACT

The literature exhibits the high power cost of RF transmissions [22], and many mechanisms have been investigated to reduce this energy budget. Some of them rely on frame size reduction such as compression [23] or data aggregation [24]. A complementary solution is to limit the number of control packets sent. For instance, RPL uses the trickle algorithm; see [25]. Several other parameters may influence the transceiver consumption, for instance throughput and packet delivery rate have direct impact on the energy consumption, as they define the duration of each transmissions and the average number of retransmission.

In our target home or building automation use case, RF nodes running over battery are periodically reporting values sensed by the probe, and downward traffic is restricted to node configuration and is not periodic. This results in highly asymmetric traffic mostly from sensors to the border router.

The RPL routing control messages practical periodicity is described in Table IV. The RF nodes mainly interact at initialization in the proactive RPL strategy. This RF node control traffic is independent from the topology. For multicast packets, a RF mode gets back to sleep just after sending whereas it waits for the acknowledgment for unicast packets. There is one exception for multicast DIS packets, where we forced the RF node's transceiver in reception mode during a duration of, by default, two seconds. This ensures that the node gets all DIO from all its potential parents. This supplementary energy cost needed to keep the transceiver active during this period is balanced by the resulting efficiency of the parent selection, which can last up to the network lifetime. Moreover, if we limit the parent selection only to the first DIO received, there is a risk that the parent communicates through a lossy link or has a high RPL rank. All this resulting in suboptimal paths, transmission retries and finally higher power consumption.

We studied our RPL's scenario traffic in order to precisely determine the communication activity. Table V shows the size and number of packets exchanges running the target monitoring application over a 24h period. The majority of the traffic is concentrated in NS/NA exchanges and data reports. The overhead induced by the routing control messages is relatively very low. Notice that NS/NA exchanges are required by our active data request mechanism, which inserts a data pending flag in the NA packets. This flag advertises that a packet intended to the RF mode is buffered by its parent. When receiving this flagged NA, the RF mode keeps its radio active until the reception of all its data packets. Moreover, these NS/NA exchanges enable a periodical check of the bidirectional connection with the selected parent. The expected transmission count, ETX, metric update accordingly to NS/NA exchanges successes and MAC retries.

TABLE IV. RPL MESSAGES PERIODICITY.

Message			Comments
Type	Origin	Periodicity	
NS	Router	Best parent 60s; Other p. 120s ; Other neighbor 360s	No NS to RF
NS	RF	Best p.: 3 to 600s	Trickle timer
NA	Router	Same as NS	
NA	RF	None	No NS to RF
DIO	Router	Imin 1s ; Imax 1050s	Trickle
DIO	RF	Not regular	Once attached inform neighbors about rank
DATA	Router, RF	600s	Application dependent
DAO	Router	DTSN increment 360s	For each DIO with incremented DTSN and if parent switching
DAO	RF	Lower period between DIS and DTSN increment	Each time DIO with incremented DTSN & if parent switching
DAO ACK	Router RF	Same as DAO	When correctly received
DIS	Router	Not regular	Only when node needs infos to be attached to DAG
DIS	RF	360 s	To request DIO from parent and update DAG infos
ACK	Router, RF	Same as unicast	For each unicast frame sent

TABLE V. SIZE AND NUMBER OF MESSAGES EXCHANGED BY A RF NODE OVER A 24H PERIOD.

Mesg type	Size (B)	Sent	Recv	Comments
NS	72	151	0	
NA	72	0	151	
DIO	111	0	24	ETX metric, DODAG conf, prefix Info
DATA	93	144	0	UDP 10 bytes payload
DAO	66	24	0	target, transit Info
DAO ACK	40	0	24	requested for each DAO
DIS	33	24	0	
ACK	5	199	344	
Mesg #		543	543	
Size (B)		28 892	17 128	

VII. RF NODE POWER CONSUMPTION

Unlike generic power consumption models like in [26], we built our simple energy model including our application. In our use case, a RF node executes periodic tasks: sleeping, waking up, running, sensing, transmitting and receiving. We

 ISBN: 978-1-61208-296-7

computed the expected energy to operate the node according to the complete scenario for different hardware components and software parameters. The sleep mode current is integrated over the time of the simulation minus the active period of time. Throughputs are the same as in Table III, and the number of packet exchanges is deducted from the message periodicity presented in Table IV. We then computed the total cost over a day according to a 4 Hz wake up frequency. We consider a RF node without any probe. Probes impact is reported independently in Section X. To complete the energy model of the radio, we add the energy spent when the transceiver stays in RX mode, waiting for acknowledgments. Notice that our simple model considers that all packets are successfully received and thus there is no retry at the application or MAC layer. Table I gives the results for our reference platform with MSP430 at 16MHz and ATRF212 against the best up-to-date components.

We first investigate the power consumption of the low dropout voltage regulator, LDO, computed from its yield characteristics, with an input of 1.5V provided by a single AAA battery and the desired output voltage of 3V.

TABLE VI. ENERGY REPARTITION OVER A DAY FOR DIFFERENT RF NODES.

Node	MSP430+ATRF212 With LDO		MSP430+ATRF212 Without		SIM3C1+SI4461 Without	
	(J)	%	(J)	%	(J)	%
LDO	18.38	82	-	-	-	-
Radio Tx	0.59	3	0.59	14	0.57	34
Radio Rx	0.20	1	0.20	5	0.14	8
CCA - Backoff - Wait for Ack	0.03	<1	0.03	1	0.04	2
MCU Wake Up	2.09	9	2.09	51	0.76	45
MCU + RF Sleep	1.22	5	1.22	29	0.19	11
Total	22.51	100	4.13	100	1.71	100
Expected Lifetime on battery (Yrs)	1.31		7.16		17.24	

Table VI shows that the LDO consumes 82% of the overall energy, which is an issue. The current drawn by the node is very low, between 10 and 20 μA. This falls in the range where the LDO efficiency is the worst, between 0,5 and 0,6, leading to a huge energy waste. In order to achieve a more efficient design, this element must be removed. A battery with the required voltage, typically 3V, is directly connected to the node. Moreover, as stated in [3], [4], the battery voltage cutoff should also correspond to the node components cutoff. The voltage regulation is not mandatory if all components used in a node can work over a voltage range matching the battery's one. Although, attention should be paid to the probe precision's behavior and to the clock drift against the voltage variation. In our reference platform, we selected a 3V battery with a capacity of 1000 mAh, a self discharge current below 1 % per year, and a dropout voltage of 2.0 V. We took into account the 1% self discharge of the battery in the model by subtracting 1% of the remaining energy at the beginning of each year.

The MSP430 requires a minimum voltage of 2.2V to run at 16 MHz. However, we observed that it is able to run correctly with a voltage supply as low as 1.8 V. Hence the entire energy from the battery will be used.

After removing the LDO on our reference platform, the major part of the power consumption is due to MSP430 wake up (51%). The sleep mode of the RF transceiver and the MSP430 represents the second bigger power consumption (29%). This is an interesting result, because most of the literature about energy consumption in WSN considers the RF transceiver as the biggest energy consumer [22]. This focus is only correct when looking at instantaneous power consumptions. However, when integrating over a long time, our results show that the transceiver is not the main consumer with a rather low periodic traffic. The expected lifetime with a 1000 mAh battery reaches 7.16 years. This matches the smart energy applications lifetime targeted by these nodes. The new ARM architecture seems very promising because it reduces the leakage current and wake up costs. This potentially increases the lifetime up to 17 years.

VIII. MCU SOFTWARE OPTIMISATION

Results from Table VI show that the most important gain may result from reducing the MCU consumption in wake up and sleep mode. While the sleep mode consumption depends on the hardware, we tried to reduce the cost of each wake up and to decrease the wake up frequency.

During each wake up, ContikiOS runs several periodic processes governed by timers that are not mandatory in our application scenario. For instance, we disabled several timers related to routing maintenance in RF nodes. We also increased several timers, such as neighbor checking, neighbor unreachability detection (NUD), route lifetime checking and I/O checking, because our application do not require fine grain timing constraints. We eventually adapted some RPL timers to the RF node mode of operation, such as the DIO and DIS timers management, and timers that governs the sleeping mode. In our building automation scenario, in order to keep an acceptable reactivity, a wake up frequency of 1 Hz is a good tradeoff in practice. We also activate the fast wake up function of the MSP that decrease the wake up time from 150 μs to only 5 μs. The corresponding expected lifetime increases from 7.16 years to 12.55 years.

The biggest part of the energy consumption is now due to the sleep mode of the components that represents 52% of the overall energy budget while the MCU wake up is only 13%.

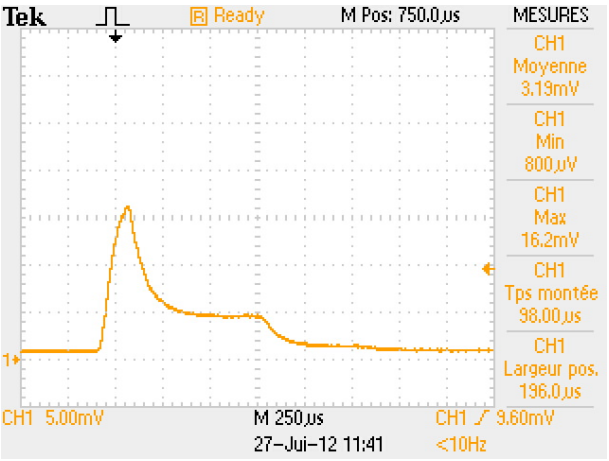
When measuring experimentally the power consumption of the MSP430 in sleep mode, we found 4.3 μA. This is far above the expected value of 1.2 μA mentioned in the data sheet when using LPM3 mode because we were using the internal oscillator (REFO). According to the data sheet this oscillator consume 3 μA. The intend of this oscillator is to provide a precise clock at 32,768 KHz. In our case, we do not need such a precision. We eventually use the internal low power oscillator (VLO) and the power consumption matched the expected value of 1,2 μA under 3 Volts. With this configuration the average current consumption of the platform drops to 5.9 μA, giving a computed expected lifetime of 20.23 years.

IX. COMMUNICATION SOFTWARE OPTIMIZATION

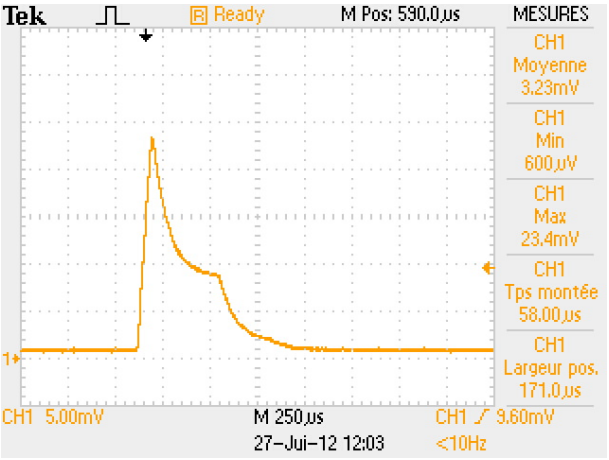
At this point, our study shows that the next optimization should focus on radio transmissions that now represents 39 % of the total energy budget.

First of all, we determined the impact of lowering the transmission power according to the transmission distance.

 ISBN: 978-1-61208-296-7



8 MHz; $H_{scale} = 250\mu s$; $V_{scale} = 5\text{mV}$



16 MHz; $H_{scale} = 250\mu s$; $V_{scale} = 5\text{mV}$

Fig. 2. Current average profile of MSP430 wake up at different frequency.

This is presented in Table III where we adjusted the 868MHz transceiver's power in order to compare with the maximum range of the 2.4GHz transceiver. The resulting energy gain is small and the 2.4GHz transceiver stays 2 order of magnitude better. Also, with a transceiver at maxim gain, the transmission reaches the wider possible range with the best PDR. The only drawback concerns the interferences but with our seldom communication application and a good MAC layer, this is not really an issue.

Hence, we focus on the limitation of the number of messages. When parsing the radio activity presented in Table V, it appears that 49.7 % of them are NS/NA, 13,6 % are RPL messages, and the remaining 36,7 % are data reports. Reducing the number of data messages is application dependent and thus not in the scope of our study. We study how to reduce RPL and NS/NA messages. Initially, NS/NA exchange happened every 10 minutes between a RF node and its parent. We augmented this interval, at the price of a greater latency for downward message transmissions, and more sparse connectivity checking. Though, notice that if we have a data reporting every 10 minutes, the link is regularly checked during these transmissions, and thus we may remove NS/NA exchanges, with some modifications of the neighbor discovery mechanism implemented in ContikiOS. In order to reuse our downward frame exchange mechanism, the data pending flag would also need to be put in the data frame's acknowledgments. Thus, data messages would also update the ETX metric and check the upward link, limiting the overhead to RPL messages only. Such an optimization of the NS/NA exchanges decreases the average node consumption leading to an expected lifetime of 28.5 years.

X. Probes Power Consumption

Embedded probes need some current for their functioning. Table VII presents the power consumption for classical WSN building monitoring probes running under 3V.

TABLE VII. Various probes energy and current under 3V.

Type	Vendor	Current (μA) active	Current (μA) standby	Duration (s)	Energy (J/yr)
Temp	TI TMP112	7	0.5	0.035	47
Temp Humid	Sensirion SHT21	300	0.15	0.114(HP) 0.015(LP)	20 15
PIR	PANASONIC EKMB1103112	1,9	1	2	95
DoorOpen	Meder KSK-1A66-1015	0.0006		NA	<1
Light	TAOS TSL2561T	240	3,2	0.25	312
CO2	AlphaSense IRC-A1	> 20000		0.46	>>1000

Embedding a CO2 probe, such as the one in Table VII, on a battery powered node is not wise since it depletes a 1000 mAh battery in only two days. On the contrary, the door/window opening probe is consuming very few energy, because its ILS bulb requires no power to operate. For other probes, there is great difference between the active and the standby consumption, hence duty cycling set up is necessary. The PIR probe cannot be fully optimized, because it automatically wakes up when it detects something, and goes back to sleep after 2 seconds. With the luminosity probe, the MCU needs to trigger a sensing window and retrieve the value acquired. As a result, the MCU can return in sleep mode during the sensing, and retrieve the luminosity value during the next wake up. The window size can be optimized as well as the periodicity of the measures. In our scenario, the sensing window period lasts 250 ms and the MCU retrieves the value during its next wake up. The temperature and temperature/humidity probes need some processing from the MCU when providing these values, increasing substantially the overall cost of the probe sensing. However, the periodicity of sensing can be large, because humidity and temperature have a low dynamic. For the SHT21, the time needed to do the measurements depends on the precision. For a full precision, 14 bits temperature, it needs a maximum time of 85 ms and for 12 bits humidity value, 29 ms. Overall, the CPU needs to be active during 114 ms. This time drops to only 15 ms for 11 bit precision temperature and 8 bit precision humidity. The TMP112 probe has a default precision of 12 bit, and an extended mode that can be activated to measure temperatures greater than +128 $^\circ$C that far exceed our requirements. According to [27], the maximum conversion time for a 12 bit temperature value is 35 ms.

Table VIII summarizes the power consumption of the probes depending on their configuration. We computed their relative part in the total energy budget, considering the node at different improvements levels following the preceding Sections. This shows that embedding a probe in a node impacts

 ISBN: 978-1-61208-296-7

TABLE VIII. POWER CONSUMPTION OF PROBES ACCORDING TO THE NODE SOFTWARE'S OPTIMIZATION (V1 SECTION VIII, V2 SECTION IX)

Probe	Precision / Period	lifetime v2 (yrs)	Energy (%)	lifetime v2 (yrs)	Energy (%)
Temp	12 bit / 10 s	4.69	77	5.03	82
Temp	12 bit / 60 s	12.43	39	15.13	47
TempHumid	Full / 10 s	1.64	90	1.73	94
TempHumid	Full / 60 s	6.43	60	7.81	73
TempHumid	Min / 10 s	7.34	54	9.20	68
TempHumid	Min / 60 s	13,18	18	20.61	28
PIR	/ 60 s	17.03	16	22.54	21
DoorOpen	/ 1 for 10 μs	20.23	< 1	28.50	< 1
Light	/ 10s 250ms	7.80	61	8.80	69
Light	/ 6 s 250ms	11.61	43	13.97	51

significantly the average power consumption. For instance, using the full precision mode of the SHT21 reduces by an order of magnitude the expected lifetime of the node, as compared to the minimum precision mode.

XI. CONCLUSION AND FUTURE WORK

In this paper, we conducted an extensive power consumption study for the design of a wireless sensor network platform. We assessed the energy consumption for the different elements of the node based on real power consumption measurements and data sheet numbers. We set up a simple consumption model in order to discuss the power optimization of the MCU, the radio transceiver and a range of probes. Moreover, we introduces software optimizations, related to our application scenario, in ContikiOS and the RPL network stack.

We pointed out the key parameters that govern the energy consumption. We implemented all the energy improvements for the selected components in order to estimate the best design for a WSN node. Our results estimate a more than 20 years RF node lifetime with a data reporting interval of 10 minutes. This encompasses our aims with a probe added within the power budget for a 10 years lifetime.

We plan to further power our nodes with an energy harvesting system such as a solar PV panel and a super capacitor. In the home or building automation applications such an autonomous sensing node is our next target design. Given the average power consumption of our actual nodes below 17μW, we are confident in the realization of such a design.

ACKNOWLEDGMENT

The authors would like to thank Mathieu Pouillot and Pierre-Emmanuel Goudet for their valuable help and support.

REFERENCES

[1] V. Raghunathan, C. Schurgers, S. Park, and M. Srivastava, "Energy-aware wireless microsensor networks," *Signal Processing Magazine, IEEE*, vol. 19, no. 2, pp. 40–50, 2002.

[2] J. McLurkin, "Algorithms for distributed sensor networks," Ph.D. dissertation, Department of Electrical Engineering and Computer Sciences, University of California, 1999.

[3] J. Hill and D. Culler, "Mica: A wireless platform for deeply embedded networks," *Micro, IEEE*, vol. 22, no. 6, pp. 12–24, 2002.

[4] J. Polastre, R. Szewczyk, and D. Culler, "Telos: enabling ultra-low power wireless research," in *IPSN*. IEEE, 2005, pp. 364–369.

[5] N. Kushalnagar, G. Montenegro, C. Schumacher, and A. Danfoss, "Ipv6 over low-power wireless personal area networks (6lowpans): Overview, assumptions, problem statement, and goals," RFC 4919, August 2007.

[6] A. Brandt, J. Buron, and G. Porcu, "Home automation routing requirements in low-power and lossy networks," RFC 5826, April 2010.

[7] T. Narten, E. Nordmark, W. Simpson, and H. Soliman, "Neighbor discovery for ip version 6 (ipv6)," RFC 4861, September 2007.

[8] J. Ko, S. Dawson-Haggerty, O. Gnawali, D. Culler, and A. Terzis, "Evaluating the performance of rpl and 6lowpan in tinyos," in *IPSN*. ACM, 2011.

[9] Q. Wang, M. Hempstead, and W. Yang, "A realistic power consumption model for wireless sensor network devices," in *SECON*. IEEE, 2006, pp. 286 – 295.

[10] C. Chauvenet, B. Tourancheau, D. Genon-Catalot, P.-E. Goudet, and M. Pouillot, "A communication stack over plc for multi physical layer ipv6 networking," in *Smart Grid Communications (SmartGridComm), 2010 First IEEE International Conference on*, 2010, pp. 250–255.

[11] L. B. Saad, C. Chauvenet, and B. Tourancheau, "Ipv6 (internet protocol version 6) heterogeneous networking infrastructure for energy efficient building," *Energy*, vol. 44, no. 1, pp. 447 – 457, 2012.

[12] C. Chauvenet, B. Tourancheau, D. Genon-Catalot, P.-E. Goudet, and M. Pouillot, "Interoperable ipv6 sensor networking over plc and rf media," *International Journal of Business Data Communications and Networking (IJBDCN)*, vol. 6, no. 4, pp. 1–20, 2010.

[13] A. Dunkels, B. Gronvall, and T. Voigt, "Contiki-a lightweight and flexible operating system for tiny networked sensors," in *29th Annual IEEE International Conference on Local Computer Networks*. IEEE, 2004, pp. 455–462.

[14] IEEE, "Ieee standard for local and metropolitan area networks–part 15.4: Low-rate wireless personal area networks (lr-wpans)," 2011.

[15] A. Dunkels, "The ContikiMAC Radio Duty Cycling Protocol," Swedish Institute of Computer Science, Tech. Rep. T2011:13, Dec. 2011.

[16] Z. Shelby and C. Chauvenet, "The ipso application framework," August 2012 (last checked June 2013). [Online]. Available: http://www.ipso-alliance.org/wp-content/media/draft-ipso-app-framework-04.pdf

[17] Z. Shelby, K. Hartke, C. Bornmann, and B. Franck, "Constrained application protocol (coap)," IETF draft, 2012.

[18] Texas Instruments, "Msp4301611 datasheet," 2010 (last checked June 2013). [Online]. Available: http://www.ti.com/lit/ds/symlink/msp430f1611.pdf

[19] ——, "Msp430f5438a datasheet," 2010 (last checked June 2013). [Online]. Available: http://www.ti.com/lit/ds/symlink/msp430f5438a.pdf

[20] Scilabs, "Cortexm3 datasheet," 2012 (last checked June 2013). [Online]. Available: http://www.silabs.com/products/mcu/mixed-signalmcu/Pages/SiM3C1xx.aspx

[21] Atmega, "Atrf212 datasheet," 2012 (last checked June 2013). [Online]. Available: http://www.atmel.com/Images/doc8168.pdf

[22] J.-P. Vasseur and A. Dunkels, *Interconnecting smart objects with ip: The next internet*. Morgan Kaufmann, 2010.

[23] C. Bornmann, "6lowpan generic compression of headers and header-like payloads," IETF draft, 2011.

[24] S. Cui, A. J. Goldsmith, and A. Bahai, "Energy-efficiency of mimo and cooperative mimo techniques in sensor networks," *Selected Areas in Communications, IEEE Journal on*, vol. 22, no. 6, pp. 1089–1098, 2004.

[25] P. A. Levis, N. Patel, D. Culler, and S. Shenker, "Trickle: A self regulating algorithm for code propagation and maintenance in wireless sensor networks," Computer Science Division, University of California, Tech. Rep., 2003.

[26] Q. Wang and W. Yang, "Energy consumption model for power management in wireless sensor networks," in *Sensor, Mesh and Ad Hoc Communications and Networks, 2007. SECON'07. 4th Annual IEEE Communications Society Conference on*. IEEE, 2007, pp. 142–151.

[27] Texas Instruments, "Tmp112 datasheet," 2009 (last checked June 2013). [Online]. Available: http://www.ti.com/lit/ds/symlink/tmp112.pdf

 ISBN: 978-1-61208-296-7

A Time-Domain Based Lossless Data Compression Technique for Wireless Wearable Biometric Devices

Chengliang Dai, Christopher Bailey
Department of Computer Science
University of York
York, United Kingdom
Email: cd633@york.ac.uk, christopher.crispin-bailey@york.ac.uk

***Abstract*—This paper presents a promising lossless compression technique called Log2 Sub-band encoding, which is suitable for implementing on wireless wearable biometric devices. Data compression promises a power saving from the transceiver during the data transmission and further extends battery lifetime on a device. The performance of this technique is measured in term of compression ratio (CR). Our simulations suggest a CR that is comparable, and indeed superior to the combination of Differential pulse-code modulation (DPCM) and Huffman coding, whilst using minimal hardware. The simulations primarily use electroencephalogram (EEG) data, and a estimated power saving is given whilst the implementation issues and possible influence of different biomedical data on technique′s performance will also be considered.**

***Keywords*—Wearable device; Bioelectric data; Lossless compression technique; Power consumption**

I. INTRODUCTION

It has been a long time since people realized that bioelectric signals could reflect alternations inside our biological system, and such signals now can help researchers to study many diseases or other external and internal stimuli that cause these alternations. As a result, wireless wearable biometric devices have drawn much attention in recent years, and they promise to make collecting bioelectric data more flexible and convenient. Applications and prototypes have been widely used in medical monitoring, brain-computer interfaces (BCI), and other relevant areas [1] [2]. Especially in medical fields, wearable devices significantly improved patients quality of living whilst allowing caregivers and doctors to provide a better healthcare for patients with various neurological and physical diseases like epilepsy, Alzheimers, and insuring early detection of emergency conditions for high risk patients.

With all the advantages of using wireless wearable devices, power consumption now becomes a design obstacle to the prevailing of these devices. A wearable device is usually battery powered, and the limited size of the device restricts the size and capacity of batteries, so power efficiency is extremely crucial for a device to extend the operational lifetime and further improve user experience. There are several ways to reduce device power consumption, and since the transceiver is one of the significant contributors to the power consumption, reducing the size of transmitted data might be a fruitful objective.

Nevertheless, compressing the data increases the complexity of the system and itself consumes power, such that it is necessary to strike a balance between the power that a compression unit is able to save and the power it consumes. A compression technique that demands less hardware resource but delivers high compression ratio (CR) will be a desirable solution, particularly where large numbers of channels become more common, and hence multiply the hardware and power costs of compression.

Bioelectric signals are highly non-Gaussian, non-stationary and non-linear, which make data compression a difficult task. Moreover, signals for clinical uses require a high consistency of original data and reconstructed data that rules out several lossy data compression techniques [3]. Various compression methods have been tested on bioelectric signals especially on EEG signals including entropy encoding [4] [5], predictive encoding [4] [6], transform-based encoding [3], etc. Entropy encoding techniques such as Huffman coding are versatile and widely used on many kinds of data, but they require the pre-knowledge of the signals, the probability density of samples for instance, before conducting any compression. This is impractical for a wireless biometric device where data are collected, processed and transmitted in real-time. Most predictive encoding techniques encounter the same problem. Transform-based encoding techniques can achieve impressive compression results, but they are either lossy or far too complex in system level. Therefore, a new method called Log2 Sub-band encoding is developed to overcome these limitations.

In the rest of this paper, the basic idea of Log2 Sub-band algorithm will be given in section II, and followed by an introduction to test data in section III and the result analysis in section IV. Preliminary hardware design will be given in section V, and finally the conclusion in section VI.

II. LOG2 SUB-BAND COMPRESSION

The fundamental design of this algorithm is simple: after analog-to-digital conversion, every data sample is divided into several bit-fields (chunks), and each chunk is compared with the same part of the previous sample. Chunks of the current sample will not be sent if they are identical to the parts of previous sample. An extra header is sent with every sample to indicate number of chunks transmitted. For instance, if a bioelectric signal is digitalized into 12 bits, the whole procedure can be illustrated with Fig. 1.

As it is shown in Fig.1, each data sample is chopped into 4-bit chunks, making 3 nibbles, and there will be four scenarios including transmitting whole sample, 2 nibbles of the sample, 1 nibble or only the header after each comparison. There will be a 2-bit header added before transmission to guide

 ISBN: 978-1-61208-296-7

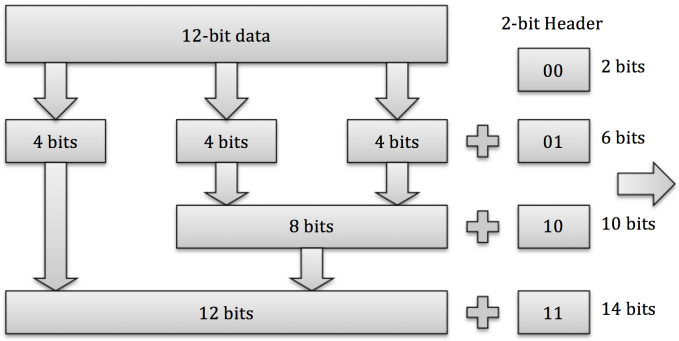


Figure. 1. Log2 Sub-band encoding

the receiver to reconstruct compressed data. The procedure is similar to the simplest predictive encoding technique, Differential Pulse-Code Modulation (DPCM) [7], which eliminates short-term redundancies by taking the difference of adjacent samples. However, the bit width of each sample after applying DPCM depends on the adjacent samples that have the biggest difference, and as for bioelectric signals in which sudden spikes may happen because of any stimuli or pathological reasons, the performance of DPCM is often disappointing. Even though such problem can be partly solved by using entropy coding unit as a second processing stage [7], it is still not compatible enough with complex bioelectric signals. The Log2 Sub-band algorithm, due to its encoding scheme, is adaptive to any sudden changes of the signal with a minimum cost of hardware, and it simply takes more bits to represent the parts with severe fluctuations and fewer bits when the signals become stable.

Meanwhile, the algorithm only processes the signal within the time-domain, and no information is dropped due to the compression technique, so the raw data will be fully reconstructed after reception, and are lossless.

III. Introduction to the Data for Testing

The simulations of Log2 Sub-band encoding are conducted with EEG signals, and they come from two sources: humans EEG data are from a research on nonlinear deterministic patterns of brain electrical signals at the University of Bonn [8], and mouse EEG data from an on-going Alzheimer disease study in Aberdeen University [9].

A. Human EEG data

All EEG signals were recorded with a 128-channel amplifier system, and digitized into 12 bit samples. The final signals were acquired at a sampling rate of 173.61 Hz, and the bandpass filter was set to 0.53-40 Hz [8].

Data were categorized into 5 groups, which are recordings from healthy people with eyes open and closed, signals originated from within and outside epilepsy diagnosed patients epileptogenic zone (seizure generating area) during the seizure free interval, and recordings of seizures. Signals from healthy people were recorded extra-cranially with severe eye movement artifacts (EOG), but EOG interference was removed manually afterwards by the research group in Bonn. Recordings of epilepsy patients were recorded intra-cranially. Each group has 100 data segments of 23.6 sec duration of signals, and 4096 samples in every segment.

B. Mice EEG data

The EEG signals of mice with novel knock-in Alzheimer were recorded intra-cranially from three areas, and then digitized into 12 bit/sample at a sampling rate of 200 Hz and 1000 Hz respectively. Each data segment contains 5 minutes of gathered EEG signals, captured using the University of York/Cybula Neural Acquisition Tracker (NAT-1) device.

IV. Result Analysis

To evaluate the performance of Log2 Sub-band encoding, the results from Huffman coding are used as a benchmark baseline. The compression ratio (CR) in the following analysis is defined as the size of original data vs compressed data such that:

$$CR = Size_{Original}/Size_{Compressed} \quad (1)$$

A. Huffman Coding

Optimal Huffman coding requires the probability distribution of signal samples to be known before compressing the data, and to achieve real time compression, a file-independent codebook was trained with 40 EEG data fragments from both healthy people and epilepsy patients in advance using the described data sets. To degrade the inter-channel correlation and minimize the size of codebook, DPCM was applied before conducting the Huffman coding, and it also renders better results than solely using Huffman coding [7]. Twenty EEG data fragments from healthy people with eyes closed and twenty more from patients epileptogenic zone were tested, and a CR of 1.82 was achieved from the former and 2.2 from the latter. The lowest CR of 1.51 was observed with unstable seizure signals. And the results are shown in Table I.

B. Log2 Sub-band encoding

As it is shown in Fig. 2, the type of data and sampling frequency have significant impacts on the performance of this technique, but in all cases the Log2 Sub-band method is superior to Huffman by between roughly an order of 5% to 15% better. Apparently, both Log2 Sub-band encoding and Huffman coding perform better when processing the signals with less fluctuations and higher similarity of adjacent samples. Huffman coding is less capable to process some bioelectric signals with a large value range since the more values they contain the longer the codes required for less frequent samples. However, Log2 Sub-band encoding is adaptive to all these signals without the need to modify the basic circuit design for each case.

Considering different results the technique delivered, it is worth looking into the changes of source signals that can be reflected by this compression technique. For instance, Fig. 3 and Fig. 4 are the distributions of length of transmitted data from healthy people and seizure periods. Healthy peoples EEG signals are relatively smooth so that most data samples were transmitted with 1 or 2 nibbles (76.5%), but there is an apparent drop to the number of 1 nibble data samples while compressing the seizure signals (31.0% to 22.0%). This

 ISBN: 978-1-61208-296-7

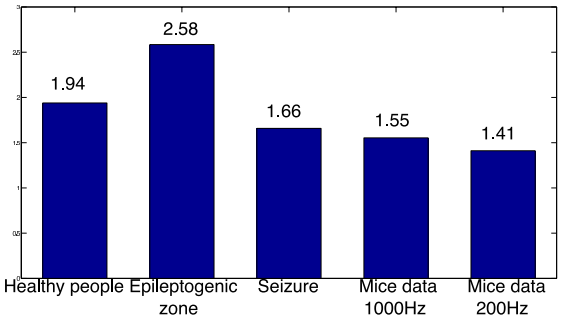


Figure. 2. Compression ratios

TABLE I
SIMULATION RESULTS

Data Type	CR of Log2 Sub-band Encoding	CR of Huffman and DPCM
Healthy people with eyes closed	1.94	1.82
Patients' epileptogenic zone	2.58	2.20
Seizure	1.66	1.51
Mice data@1000Hz	1.55	N/A
Mice data@200Hz	1.41	N/A

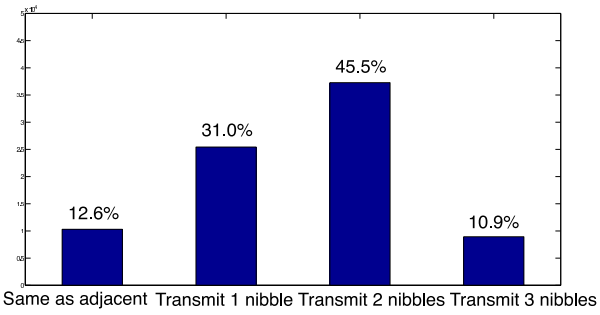


Figure. 3. The distribution of the length of transmitted healthy peoples data samples

noticeable discrepancy could be used to warn the caregivers that patients physical conditions might have changed. And this preliminary pattern recognition will not add on more workload to the system.

To reduce the power consumption is the primary purpose of introducing this encoding method. An off-the- shelf Bluetooth transceiver nRF8001 [10] is chosen to evaluate its effect. The power of this transceiver is 22mW, and it takes 22nJ to transmit 1 bit data. Considering the lowest CR (1.41) reached earlier, which is around 30% of size reduction, and if uncompressed data were supposed to be transmitted at 64Kb/s, this encoding method would bring at least 200μW's power saving before introducing the power overhead of the compression itself. Hence, as long as the compression circuit consumes less than 200μW a significant saving might be obtained for each recorded channel.

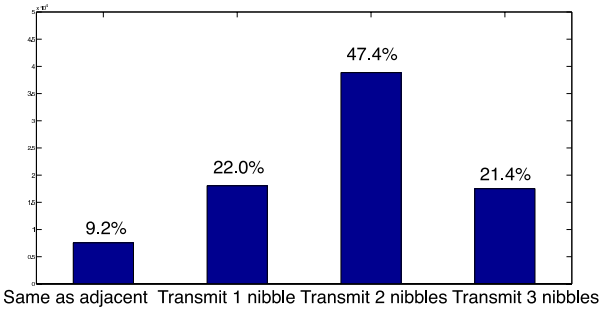


Figure. 4. The distribution of the length of transmitted seizures data samples

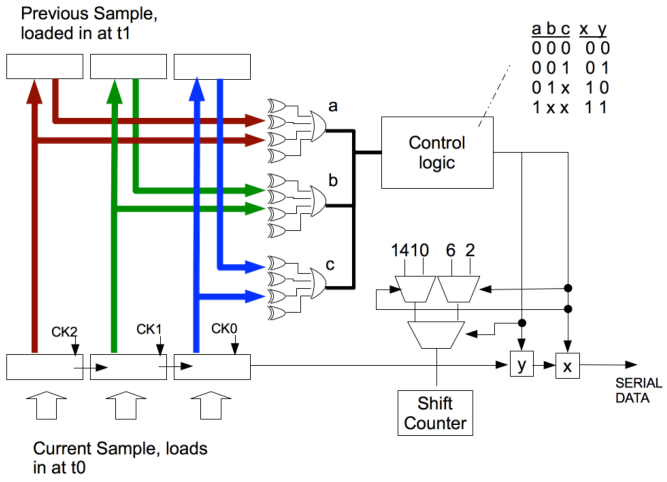


Figure. 5. Simplified Hardware Design

V. HARDWARE IMPLEMENTATION

A simplified hardware design of this encoding method for processing the EEG signals above mentioned is shown in Fig. 5. Current sample and previous sample are loaded respectively, and different sub-bands (chunks) are sent to logic gates to compare, with the result used to determine the header and to control a shifter that shifts right number of bits to transmit. The number of bits represents the data payload of 0,4,8, or 12 bits, and the header of 2 bits in each case. In this case, it would therefore be a total symbol length of 2 bits, 6 bits, 10 bits, and 14 bits.

VI. CONCLUSION

As mentioned earlier, Log2 Sub-band encoding has a better performance on EEG data and simpler hardware design compare with other current techniques, based upon our initial evaluations. Since bioelectric signals share many common behaviors, we can anticipate that the Log2 Sub-Band method can be applied widely to other signals besides EEG, and a more diverse analysis would be very desirable. Improvements to the system may be possible. We have already considered the possibility of variable-sub-band encoding, such that the width of bands can be varied in such a way as to ensure that the distribution of changes in each band is optimized, for instance {3,3,4,6} might be better for some signal types with a resolution of 16-bit. Meanwhile, the possibility of switching

 ISBN: 978-1-61208-296-7

between two encodings depending upon the relative activity or idleness of signal behavior might permit higher overall compression rates with relatively minimal circuit changes.

ACKNOWLEDGMENT

We are grateful for access to the mouse EEG data supplied by the Institute of Medical Sciences, Universysity of Aberdeen.

Mouse data was recorded using the NAT-1 device, developed by University of York and Cybula Ltd. This device was developed with support from the UK Technology Strategy Board.

REFERENCES

[1] L. Boquete et al., "A portable wireless biometric multi-channel system," Measurement, vol. 45, no. 6, pp. 1587–1598, Jul. 2012.

[2] U. Hoffmann, J.-M. Vesin, T. Ebrahimi, and K. Diserens, "An efficient P300-based brain-computer interface for disabled subjects." Journal of neuroscience methods, vol. 167, no. 1, pp. 115–25, Jan. 2008.

[3] W. Optimization, L. Brechet, M.-f. Lucas, C. Doncarli, and D. Farina, "Compression of Biomedical Signals With Mother Wavelet Packet Selection," IEEE transactions on bio-medical engineering, vol. 54, no. 12, pp. 2186–2192, 2007.

[4] G. Antoniol and P. Tonella, "EEG data compression techniques." IEEE transactions on bio-medical engineering, vol. 44, no. 2, pp. 105–14, Feb. 1997.

[5] D. O'Shea and R. McSweeney, "Efficient Implementation of Arithmetic Compression for EEG," IET Proceedings of Irish, 2011.

[6] N. Sriraam, "A high-performance lossless compression scheme for EEG signals using wavelet transform and neural network predictors." International journal of telemedicine and applications, vol. 2012, p. 302581, Jan. 2012.

[7] Y. Wongsawat, S. Oraintara, T. Tanaka, and K. Rao, "Lossless Multi-channel EEG Compression," 2006 IEEE International Symposium on Circuits and Systems, vol. 6, pp. 1611–1614, 2006.

[8] R. Andrzejak, K. Lehnertz, F. Mormann, C. Rieke, P. David, and C. Elger, "Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state," Physical Review E, vol. 64, no. 6, p. 061907, Nov. 2001.

[9] B. Platt et al., "Abnormal cognition, sleep, EEG and brain metabolism in a novel knock-in Alzheimer mouse, PLB1." PloS one, vol. 6, no. 11, p. e27068, Jan. 2011.

[10] "nRF8001 Bluetooth® low energy Connectivity IC," http://www.nordicsemi.com/eng/Products/Bluetooth-R-low-energy/nRF8001, 2012, [Online; accessed 09-April-2013].

 ISBN: 978-1-61208-296-7

Data Fusion in Wireless Sensor Networks using Fuzzy Set Theory

Ali Berrached and Andre de Korvin
Department of Computer and Mathematical Sciences
University of Houston-Downtown
Houston, Texas 77002 USA
berracheda@uhd.edu dekorvina@uhd.edu

***Abstract*—Wireless Sensor Networks (WSNs) are collections of sensor nodes deployed in a geographical area with the purpose of monitoring the environment in which they are deployed and detect events of interest. Sensor nodes are tiny devices with limited battery power and communication range. Data fusion is the process of combining raw data from the various sensor nodes to obtain information of greater quality and make accurate decisions about the events of interest. Given that energy is the main constrain in WSNs, data fusion can also be used to reduce the volume of data transmitted over the network, thereby extending the network's lifetime. In this paper, we propose a data fusion framework that uses fuzzy set theory to aggregate data from multiple sensors at the cluster level. The algorithm is capable of handling the inherent inaccuracy and conflicts in environmental data readings.**

Keywords-wireless sensor networks; data fusion; data aggregation.

I. INTRODUCTION

Wireless Sensor Networks (WSNs) are collections of large numbers of sensor nodes capable of collecting, relaying, and processing sensor readings from the physical world. They have a wide range of applications in both military and civilian environments ranging from natural habitat monitoring to enemy detection and tracking in the battlefield [1]. In most cases, power sources in the sensor nodes are not rechargeable- they are battery based and the nodes are deployed in remote and/or hostile environments. Since a sensor network is usually expected to operate for several months without recharging, energy conservation is an important design objective to prolong the network's lifetime [2]. At the same time, the network's ability to collect and communicate the data of interest on a timely fashion is also a critical objective.

In most applications, sensor nodes are deployed randomly and in large numbers over a target area (e.g., dropped from an airplane). Since data is collected by a large array of densely deployed neighboring nodes, there tends to be a high degree of redundancy and correlation in the data collected. Additionally, due to the harsh conditions in which sensors are often deployed, they tend to be prone to various source of errors such as noise from external sources, hardware noise, sensors inaccuracies and imprecision, and various environmental effects [3]. Data fusion is the process of combining raw data from various sensor nodes. Its main purpose is to obtain information of greater quality and make accurate decisions about the events of interest based on the data collected from the various sensors [4][7][8]. Our proposed data fusion framework uses Fuzzy set theory [6] to handle the inaccuracies that are inherent in sensor data readings and for combining conflicting information from various sources. Moreover, by performing data aggregation at the cluster-level, an additional benefit is the reduction of the amount of raw data transmitted over the network, which prolongs the network lifetime. A secondary benefit is to prevent flooding the Base Station with raw data.

In the next section, we lay out our data fusion framework and its underlying assumptions. In Section 3, we present a simple cluster-level data fusion algorithm that assumes accurate sensor readings. In Section 4, we present a base station-level aggregation algorithm. In Section 5, we present our cluster-level fuzzy fusion algorithm, and in Section 6, we present our concluding remarks and future extensions.

II. DATA FUSION FRAMEWORK

We assume a network of sensors deployed randomly and in large numbers in a bounded area where events of interest are expected to occur. Sensors are assumed to be stationary (i.e., not mobile). They are capable of self-organizing into clusters with one node serving as a cluster head for each cluster. The clustering algorithm partitions the network into groups of nodes each covering a subset of the whole coverage region. All nodes in a cluster send their data readings to their cluster head which aggregates and forwards its

 ISBN: 978-1-61208-296-7

decisions to the base station. The base station collects information coming from all cluster heads and produces the final decisions. Figure 1 depicts a diagram of a cluster based wireless sensor network.

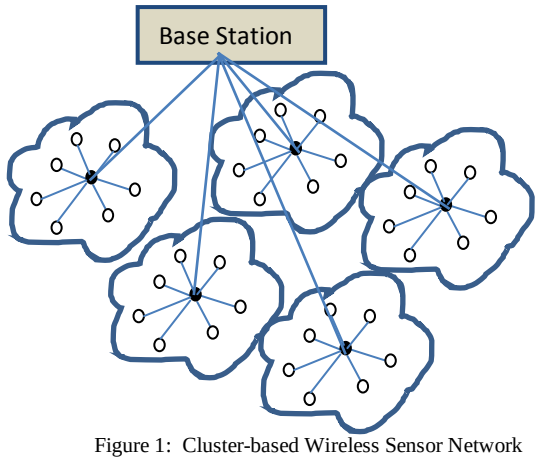


Figure 1: Cluster-based Wireless Sensor Network

There are several clustering algorithms proposed for WSNs [5]. We assume that each cluster consists of N sensor nodes, with N possibly different for each cluster, and each node is capable of sensing one feature or attribute of interest such as the surrounding temperature, humidity, light intensity, smoke density, amount of Carbone Monoxide (or any other chemical), etc.

We propose a framework whereby a *data fusion processor* is employed at the cluster-level. The *data fusion processor* serves to aggregate the raw data transmitted by the cluster nodes and generate decisions that are transmitted to the base station.

III. CLUSTER-LEVEL DATA FUSION

In each cluster, we assume one sensor is activated to monitor one of the features of interest. Assuming there are *m* features to be monitored, there will be *m* sensors with the i^{th} sensor observing feature F_i. The *m* readings are to be aggregated by the cluster *data fusion processor* to reach a decision concerning the occurrence of an event of interest (e.g., the intrusion of an enemy or the occurrence of a fire).

We assume each cluster fusion processor is provided with a local decision matrix **D** defined as:

$$\mathbf{D} = [D_1, D_2, \ldots, D_n]$$

where D_k is a vector of feature values ($[f_{1,k}, f_{2,k}, \ldots, f_{m,k}]$) supporting decision d_k.

As an example, let us consider an application where a sensor network is used for the detection of fire in a forest. The set of monitored features may consist of the following:

F_1: temperature
F_2: smoke density
F_3: humidity

The set decisions:

d_1: Fire unlikely
d_2: Fire likely

A decision matrix can be defined as:

$$\begin{bmatrix} \llcorner 70 & 130\urcorner \\ \llcorner 10 & 70\urcorner \\ 70\urcorner & \llcorner 20 \end{bmatrix}$$

where $\llcorner x$ indicates any value smaller than x is replaced with x and $x\urcorner$ indicates any value larger than x is replaced with x. The above example decision matrix indicates that if feature F_1 (temperature) is 70 (degrees Fahrenheit) or less, feature F_2 (smoke density) is 10 or less, and F_3 (humidity) is greater or equal to 70 then decision d_1 is taken (i.e., Fire unlikely); and if F_1 (temperature) is 130 (degrees Fahrenheit) or more, F_2 (smoke density) is 70 or more, and F_3 (humidity) is 20 or less then decision d_2 is taken (i.e., Fire likely).

Given an actual sampling/reading $\mathbf{R} = [r_1, r_2, \ldots, r_m]$ collected by a cluster from its *m* sensors, the cluster *data fusion processor* takes decision d_j so that:

$$\sum_{i=1}^{m}(f_{i,j} - r_i)^2) \leq \sum_{i=1}^{m}(f_{i,k} - r_i)^2 \quad (1)$$

for all k = 1, 2, …n.

Note that $\sum_{i=1}^{m}(f_{i,k} - r_i)^2$ is a measure of how close the actual data collected by the sensors is to the feature values expected to support decision d_k. Conversely, we can define the strength of a decision d_k as the inverse of the Cartesian distance:

$$Strength(d_k) = \frac{1}{\sum_{i=1}^{m}(f_{i,k} - r_i)^2)} \quad (2)$$

The larger the distance $\sum_{i=1}^{m}(f_{i,k} - r_i)^2$ of the data reading **R** to the feature values expected to supported decision d_k, the weaker the decision.

For the example decision matrix above, if the data reading R = [140, 50, 30] then:

Strength(d_1) = 1/($(70-140)^2$ +$(10-50)^2$ + $(70 - 30)^2$) = 1/8100

Strength(d_2) = 1/($(130-130)^2$ +$(70-50)^2$ + $(20 - 30)^2$) = 1/500

Decision d_2 (*Fire like*ly) has more strength which, intuitively, is the expected decision for the example reading.

IV. BASE STATION-LEVEL DECISION FUSION

Assuming there are S clusters (where S can vary from cluster to cluster), the base station receives S decisions d_{j_1}, d_{j_2}, …d_{j_s}, one from each cluster head. If the clusters are disjoint, then each cluster decision is representative of the data collected from the area covered by its sensors. On the other hand, if the clusters have overlapping coverage then a number of aggregation algorithms can be used to select a decision from the S decisions received. The simplest method would be to use a simple voting scheme, where the selected decision is the one generated by the most clusters. A better approach would be to use the strength of each decision in the selection process, where the strength of a decision is defined in equation (2). The disadvantage of this approach is that each cluster head needs to transmit, in addition to its decision, the decision strength of its decision. It can, however, lead to more accurate decisions. The additional information can, for example, be used to weed out weak decisions, by considering only decisions that have decision strength above a certain threshold. This can reduce any bias caused by clusters that don't have enough information to make a strong decision (because they are far from the event of interest, for example). From the remaining decisions, we can select the decision that has the majority of votes, or the one with the largest average strength.

V. FUZZY DATA FUSION

The data fusion process described above assumes exactly one sensor is deployed to monitor each feature and sensor readings are accurate and precise. Most often, however, environmental data tends to be vague and noisy, and sensors tend to be prone to errors and malfunction. To overcome these limitations, multiple sensors can be activated to monitor each feature of interest. Fuzzy set theory allows us to map inaccurate crisp sensor readings into fuzzy values that include a measure of confidence or belief of the accuracy of the readings. It also allows us to aggregate conflicting data readings. We first give a brief overview of fuzzy sets.

A. *Overview of Fuzzy Sets*

A fuzzy set A on space X is defined by its membership function:

$$A: X \rightarrow [0, 1]$$

The membership function is a generalization of the characteristic function of a crisp (i.e., ordinary) set. For each $x \in X$, A(x) denotes the degree to which element x is a member of fuzzy set A. For a crisp set, of course,

$$A(x) = \begin{cases} 1 \; iff \; x \in A \\ 0 \; Otherwise \end{cases}$$

For fuzzy sets, $0 \leq A(x) \leq 1$. Those x's for which $A(x) > 0$ constitute the support of fuzzy set A. For notational convenience, we do not distinguish between the membership function and the fuzzy set itself. In effect, the membership function is the fuzzy set. When the domain $X = \{x_1, x_2, \ldots, x_n\}$ is finite, we represent fuzzy set A by the notation:

$$A = \sum_{i=1}^{n} \mu_i / x_i$$

where $\mu_i = A(x_i)$ denotes the degree to which x_i belongs to *A* or the confidence of the belief that x_i belongs to *A*

B. *Data Fuzzification*

A stated above, we assume within each cluster, each feature is observed by an array of *s* sensors. Therefore, for each of *m* features F_i, *s* readings are collected by each cluster. The process of data fuzzification begins by partitioning the domain of each feature F_i into a set of *v* intervals, I_1, I_2, …, I_v and mapping each of the *s* readings into the interval that it belongs to. For each interval I_k, we can get the fraction α_{ik} of sensors observing feature F_i whose reading falls in interval I_k.

If we represent each interval I_k with its mid-point P_k, we obtain a function:

$$f_i: P_k \rightarrow \alpha_{ik}, \text{ for all } k = 1, 2, \ldots, v$$

where *v* is the number of sampling intervals and *i* is the i^{th} feature. After normalizing f_i and taking a linear interpolation, we obtain a continuous function $\bar{f}_i$ that peaks at 1. The normalized values α_{ik} for feature F_i

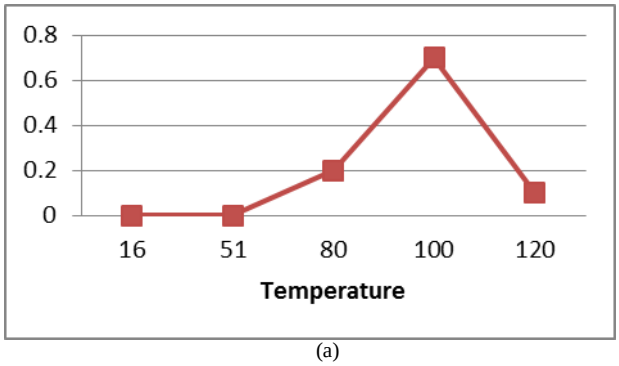
 ISBN: 978-1-61208-296-7

represent the confidence level that the value of feature F_i is P_k.

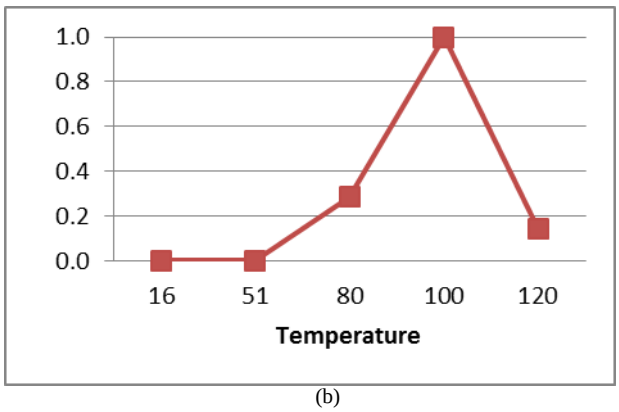
$$\hat{f}_i : x \in X \rightarrow \mu_x, \text{ where } \mu_x \in [0, 1]$$

As an example, let us assume a given cluster has 10 temperature sensors with the following readings: 97, 95, 78, 99, 102, 98, 97, 82, 119, 96. Let us also assume we have 5 sampling intervals I_1=[0, 32[, I_2=[32,70[, I_3=[70, 90[, I_4=[90, 110[, and I_5=[110, 130]. Alternatively, we can express these temperature intervals in linguist terms such very low, low, medium, high, and very high.

Figure 2(a) shows the fraction of readings for each of the sampling intervals represented with the interval's midpoint and Figure 2(b) shows its normalized counterpart.



(a)



(b)

Figure 2: Example data fuzzification

Based on Figure 2(b), the temperature value deduced from the 10 sensor readings is

$$temperature = 0.28/80 + 1/100 + 0.14/120$$

indicating a high confidence that the temperature is close to 100 and lower confidence for the other two values.

C. Cluster-Level Decision Making

As in the crisp case, each cluster head uses a decision matrix:

$$\mathbf{D} = [\mathbf{D}_1, \mathbf{D}_2, \ldots, \mathbf{D}_n]$$

where $\mathbf{D}_l$ is a vector $[f_{1,l}, f_{2,l}, \ldots, f_{m,l}]$ of the ideal feature values supporting decision d_l. Ideally, if the feature values computed by a cluster head (based on its senor readings) are exactly equal to vector D_l, then the cluster head (i.e., the data fusion processor) should take decision d_l. However, in general the computed value will not match exactly any of the vectors D_l. The objective of the decision making process is to select the decision that matches best the computed features values.

Let $\mathbf{R} = [\hat{r}_1, \hat{r}_2, \ldots, \hat{r}_m]$ be a set of fuzzy membership functions computed by a cluster head based on its sensor readings, where $\hat{r}_i$ is the membership function for feature F_i. Note that each $\hat{r}_i$ is a fuzzy membership function computed using the procedure described in the previous section. Using the notation of the previous section:

$$\hat{r}_i = \sum_{k=1}^{n} \mu_{ik} / x_{ik}$$

where $\mu_{ik} \in [0, 1]$ for all i and k.

We define the strength of decision d_l as:

$$\text{Strength}(d_l) = \sum_{i=1}^{m} min\left(\sum_{i=1}^{k} max\left(\mu_{ik}\, e^{-(x_{ik}-f_{i,l})^2}\right)\right) \quad (3)$$

Noting that the *min* function produces the weakest link among a set of series links and the *max* function generates the strongest link among a set of parallel links, we use the *max* function to weed out readings that either have low confidence level or large distance to the ideal value for each feature, then we use the *min* function to represent the strength of a decision with the strength of its weakest feature reading. Other aggregate functions such as the mean can also be used instead of the *min* and the *max* functions.

As an example, let us assume the decision matrix below with 2 features F_1 and F_2 and 3 decisions d_1, d_2, and d_3:

$$\begin{matrix} d1 & d2 & d3 \end{matrix}$$
$$\begin{bmatrix} 1 & 3 & 5 \\ 2 & 4 & 6 \end{bmatrix}$$

 ISBN: 978-1-61208-296-7

Assume at a given cluster, we get the following feature membership values based on our sensor readings and using the fuzzification procedure of the previous section:

$$\hat{r}_1 = 0.8/1 + 0.2/3$$
$$\hat{r}_2 = 0.4/3 + 0.6/4$$

Intuitively, the readings support decisions d_1 and d_2, while decision d_3 should be the weakest. Applying equation (3), we get the strength of each decision:

Strength(d_1) = min[(max($0.8\, e^{-(1-1)^2}$, $0.2\, e^{-(3-1)^2}$), max($0.4\, e^{-(3-2)^2}$, $0.6\, e^{-(4-2)^2}$)]

Strength(d_1) = min(max(0.8, 0.0036), max(0.268 , 0.011)) = 0.268

Similarly, we compute
Strength(d_2) = min(max($0.8\, e^{-(1-3)^2}$, $0.2\, e^{-(3-3)^2}$), max($0.4\, e^{-(3-4)^2}$, $0.6\, e^{-(4-4)^2}$))

Therefore, Strength(d_2) = min(0.2 , 0.6) = 0.2

Strength(d_3) = min(max($0.8\, e^{-(1-5)^2}$, $0.2\, e^{-(3-5)^2}$), max($0.4\, e^{-(3-6)^2}$, $0.6\, e^{-(4-6)^2}$))

Therefore, Strength(d_3)= min(0.00366, 0.011) = 0.00366.

Based on the above results, decision d_3 is the weakest, which intuitively is what we expected, while decision d_1 and d_2 are stronger, with d_1 slightly stronger than d_2.

VI. CONCLUSION AND FUTURE WORK

Data fusion in wireless sensor networks is critical to reduce raw data transmission across the network, which would in turn increase the network's lifetime and prevent flooding the base station. In this paper, we presented a cluster-level data fusion algorithm based on fuzzy set theory that's capable of handling inaccurate and conflicting sensor readings. We plan to extend this algorithm to aggregate sensor readings over time so that sensor nodes update their cluster heads only when significant changes occur in the sensed feature readings, and in turn, cluster heads update the base station only when a significant change occurs in the decision taken or it confidence level. We also plan to further investigate decision fusion among partially overlapping clusters.

REFERENCES

[1] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cyirci, "Wireless sensor networks: A survey," Computer Networks 38(4), 2002, pp. 393–422.

[2] P. Berman, G. Calinescu, C. Shah, and A. Zelikovsky, "Efficient energy management in sensor networks," Ad Hoc and Sensor Networks, 2005, pp. 71-90

[3] E. Elnahrawy and N. Badri, "Cleaning and querying noisy sensors," Proceedings of the 2nd ACM international conference on Wireless sensor networks and applications, 2003, pp. 78-87.

[4] R. C. Lou, C. Yih, and K. L. Su, "Multisensor fusion and integration: Approaches, applications, and future research directions," IEEE Sensors J. 2, 2 (April 2002), pp. 107–119.

[5] J. C. Maxwell, "A Treatise on Electricity and Magnetism," 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp. 68–73.

[6] W. Pedrycz and F. Gomide, "An Introduction to Fuzzy Sets Analysis and Design," MIT Press, Cambridge, Massachusetts, 1998.

[7] L. WALD, "Some terms of reference in data fusion," IEEE Trans. Geosci. Remote Sens., 13, 3 (May 1999), pp. 1190–1193.

[8] N. Xiong and P. Svensson "Multi-sensor management for information fusion: issues and approaches," Information Fusion, 2002, vol. 3, pp. 163–186.

 ISBN: 978-1-61208-296-7

An Energy Consumption Model for a Wireless Sensor Network Node based on the division of the Duty Cycle

José. M. Alcalá, and Álvaro Hernández
Department of Electronic
University of Alcalá
Alcalá de Henares, Spain
jmanuel.alcala@depeca.uah.es

Victor Cionca, Michael Hayes, and Brendan O'Flynn
WSN Group
Tyndall National Institute
Cork, Ireland
victor.cionca@tyndall.ie

Abstract— **As the energy consumption in an energy harvesting wireless sensor network is constrained, applications that can run on the nodes are limited as well. Thus, it is not only important to know how much energy can provide a harvesting system, but also how much energy will consume a specific application running on a node, and match both models to achieve a functional system from an energy point of view. Much literature can be found regarding energy harvesting models to predict energy produced but little about how this energy will be used. This paper presents an energy consumption model for nodes in a Wireless Sensor Network (WSN), which depends solely on the Duty Cycle. Likewise, applications are characterized for their duty cycle and this parameter will determine if a particular application can run on the node. The novelty of this model is the division of the duty cycle into smaller duty cycles for each of the WSN application components, allowing more control over the energy consumption. To the best of our knowledge, no other related studies have been carried out so far and we believe that this approach will contribute to develop applications for energy constrained system according to a maximum duty cycle. This approach has been tested against real data showing a relative error up to 3.14%.**

Keywords-wireless sensor networks; energy consumption model; energy harvesting; efficient energy; duty cycle; WSN applications

I. INTRODUCTION

The deployment of large Wireless Sensor Networks (WSN) provides greater granularity over different parameters (temperature, humidity, light intensity, Carbon diOxide (CO_2) levels, occupancy, etc.) in commercial and residential buildings. Consequently, this increases the control and monitoring of systems such as Heating, Ventilation, and Air Condition (HVAC), lights, doors and windows and so on. However, the maintenance of WSN can be a tedious process as batteries need to be replaced and, in most cases, sensor devices are located in places with difficult access. Hence, energy harvesting methods are widespread studied to provide long-term energy to autonomously power a mesh sensor network with a large number of nodes.

Energy harvesters can use sources from within the building such as indoor lights and vibrations, but all these sources generate relatively low and variable power, which is difficult to use. Developing low power hardware and protocols is no longer enough. The node must be aware of the amount of energy it consumes to be able to plan future activities against predicted levels of generated energy. For instance, having indoor lights as source, the power management must predict that, at night, lights are off and, consequently, reduce network activities to the minimum during this period to keep the nodes alive. Because the duty cycle determines the network activity, the energy consumption in a WSN node has a tight relation to this parameter.

Our contribution is an approach to an Energy Consumption Model based on an in-depth analysis of the duty cycle. Instead of considering a black-box model where a unique node duty cycle determines the energy consumption, our model splits the duty cycle into smaller duty cycles where each of them is attached to a specific activity performed by the node. Thus, the discretization of the duty cycle allows a greater granularity over the energy consumption and the customization of the activities depending on the available and expected energy.

To validate the model, several tests have been carried out, divided into two different experiments: firstly, an energy consumption model for a single node communicating directly to the base station has been developed and evaluated; then, the model has been adapted for a node as part of a WSN. Devices used for these tests are Tyndall motes transmitting on 2.4 GHz running TinyOS and using the Collection Tree Protocol (CTP).

In the next section, related work will be presented, followed by the experiments and modeling of the energy consumption in a node; two cases are considered: in section 3, a node communicating directly to the base station and in section 4, a forwarding node in a WSN. Finally, a discussion about the model is given in section 5, with future work and conclusions closes the paper.

II. RELATED WORK

Reducing the power consumption of wireless sensors or extending their lifetime has always been at the focus of the research community. This is handled across the levels of the platform, with low-power hardware, efficient Medium Access Control (MAC) protocols [1] and [2] that minimise the amount of time the radio is kept active, routing protocols [3] that attempt to even out the energy consumption throughout the network or data aggregation techniques [4] that reduce the overall amount of network traffic.

Energy harvesting technology [5] adds a new dimension to power management: the variability of the power source. This must be efficiently used by updating the system

 ISBN: 978-1-61208-296-7

performance: increased during high energy periods and reduced otherwise. There are many approaches in literature that address the control of system duty cycle based on available power. Possibly the most comprehensive is by Kansal, Hsu, Zahedi and Srivastava [6], who model both the energy input and the behaviour of the node as a duty cycled load. The power supply is modelled as a bursty energy source and a storage element with leakage current. In an older paper, Kansal and Srivastava [7] propose a power management framework that could perform task scheduling based on the energy input and usage model. However, it is unclear whether the complete framework was implemented and validated. Vigorito, Ganesan and Barto [8] also address the problem of managing the duty cycle of the node, but relying only on past values for harvested energy. Their algorithm is based on adaptive control theory.

In [9], Fan, Zheng and Sinha approach the problem of data rate management in energy harvesting powered multi-hop networks. They present algorithms to determine data rates for each node, based on expected traffic flow. The algorithm works best in static topologies.

The ideal is being able to predict the power consumption of a node, based on its hardware and firmware. This would allow network deployers to determine how long the node will function, and if the lifetime is not satisfactory, to update parameters in the firmware in order to extend it. There are many power consumption models that follow closely the hardware platform (power consumption per clock cycle) [10] and physical radio layer [11]. PixieOS [12] is a WSN OS that uses resource-aware programming in order to increase the efficiency.

All of the papers enumerated here consider the node a single duty cycle load, and discuss how that duty cycle can be modified. A WSN application has many intrinsic duty cycles and changing the overall duty cycle is non-trivial. One could impose a specific duty cycle at low levels in the software stack but it would affect the functionality of not only the node, but the rest of the network. This paper proposes a power consumption model that takes into account the duty cycles of the power hungry firmware components. By changing those duty cycles, the power consumption will be modified and the functionality of the node and network will not be compromised.

III. ENERGY CONSUMPTION IN A SINGLE NODE

As a first approach, the model was tested in a simple scenario where an isolated node communicates directly to the base station without forwarding any packets. Therefore, the application has less components and less complexity. Further, developing this model allows isolating the node's own activities from the activities of maintaining the topology and routing packets.

A. Energy Consumption Model

In a scenario where the node communicates directly to the base station, "send" and "sense" activities are the main tasks performed. Consequently, the energy consumption can be expressed as follows:

$$E = E_{sense} \times D_{sense} + E_{tx} \times D_{tx} + E_{sleep} \times (1 - D_{sense} - D_{tx}) \quad (1)$$

The total energy consumption for a single node can be estimated by the sum of the different components of the application and these components are determined by their duty cycle. Thus, E_{sense} is the energy consumed every time the node is sensing and it is empirically determined as well as E_{tx}. D_{sense} is the duty cycle for sensing, *tx* refers to the send component and *sleep* to the sleep modes of the node's microcontroller.

This model was tested against real data and results are presented in the following section.

B. Validation of the model

For testing the model, a single node is set to communicate directly to a base station. The application running on the node senses light intensity with a T_{sense} period and transmits directly to the base station an average of each S samples. A power analyser is used to measure the real power consumption of the node.

Three different tests have been carried out, lasting ten minutes each, to evaluate the model for different values of the sense and send duty cycles, as follows:

- Test 1: D_{sense}=0. 34% and D_{tx}=0. 19%
- Test 2: D_{sense}=2.75% and D_{tx}=0. 39%
- Test 3: D_{sense}=0.017% and D_{tx}=0.002%

In Test 3, the activity is very low so it will show how the model fits when the sleep mode is dominant.

A simulation of the model was done to compare it to the real data. In Figure 1, the evolution of the energy consumption shows that the model fits quite well with the real data. There are two main sources of error: CPU activity for handling timer overflows – these are the smallest spikes in the real data plot; and a hardware irregularity causing a slight "bump" in the current consumption right after sending a message. Both could be easily integrated into the model but they are platform specific and it is more valuable to achieve a general model. The accumulative energy consumption plots in Figure 2 shows that the real energy consumption is linear and it can therefore be modeled using a linear equation as shown in (1), where the duty cycle determines the slope of the figure. The difference between slopes of the

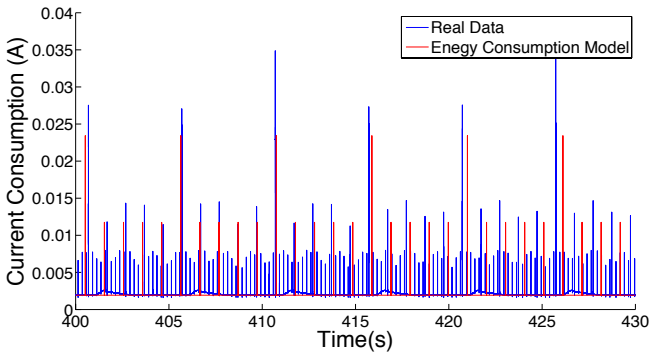


Figure 1. Simulation for Test 1. D_{sense}=0.0034 and D_{tx}=0.0019

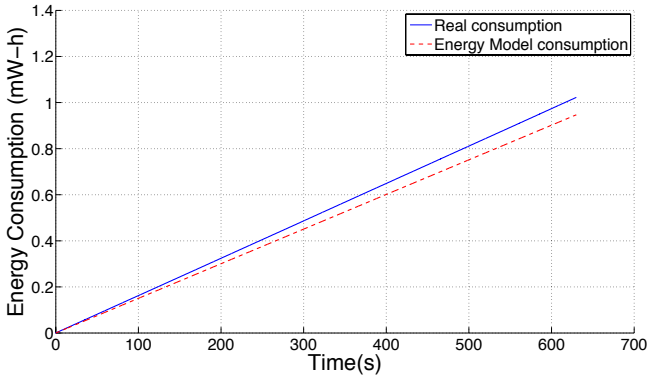


Figure 2. Acumulative Energy Consumption for Test 1.

real data and the model is due to the two error sources indicated above. An analysis of the total energy consumption based on duty cycles is shown in Figure 3. For each test, the total real consumption is compared to the energy consumption model, which is divided in the consumption of the different application components. Thus, it can be observed that the main consumption is due to the sleep mode, which is not the common case on low-power devices but, in this case, as the goal is to obtain an accurate model of the energy consumption and not a low-power system, this aspect was not taken into account. The sleep consumption should be a known constant, which can be modelled with a minimum error.

The model seems to fits perfectly on Test 3, as duty cycles for sensing and sending are negligible and almost all the consumption is due to the sleep mode. For test 2 and 3, it can be observed how the energy consumption for sending and sensing proportionally increase as their duty cycles increase as well. Especially test 2 shows the importance of the analysis of the duty cycle that the model carries out, although the energy consumption for a single sending activitiy is 2 times higher than a sensing activity, the overall is a higher energy consumption on the sensing mode as the duty cycle is 7 times bigger than the duty cycle for the sending mode. The relative error is shown on top of the bars.

IV. ENERGY CONSUMPTION IN A WSN NODE

For a node in a Wireless Sensor Network, the previous model does not fit with the real consumption because there are other important activities to take into account: maintaining the topology of the network and forwarding network traffic. In this section, the energy consumption model for a WSN node will be presented, considering the main activities by their duty cycles and then, the model will be tested against the consumption of a real node in a network.

A. Energy Consumption Model

A generic application in a WSN node performs the following major activity cycles:

- Periodically checking for incoming packets at the MAC layer: switch on the radio, check for incoming packets, switch off the radio.
- Sending **periodic beacons** at the routing layer to maintain the network topology.
- **Forwarding packets** of upstream neighbours towards the sink, at the routing layer
- Performing its own periodic activities: sensing, sending packets, etc.

From these, the MAC cycle, routing beacon and application activity cycle are predetermined, hard-coded in the node's firmware. Therefore they can be controlled. The only one that cannot be controlled is the rate of packet forwarding, as this depends on the network dynamics.

In a static WSN with a static routing tree, the packet forwarding rate could be determined for each node, based on the structure of the network sub-tree routed at that node and the activity cycle of the application components, as shown in (2) where the number of forwarded packets is proportional to the time (t) (with t_o as the initial time) and T_{app} and N_{sons} are the period of the application and the number of sons for that node, respectively. Therefore, in a static WSN it should be possible to provide an accurate estimate of the network´s energy consumption depending on the different components of the application, as it shows in the (3). E_{mac}, $E_{beacons}$ and E_{app} are predetermined and empirically obtained and stand for the energy of the periodic MAC check, periodic beacons and forwarding packets of a duty cycle; D_x stands for the duty cycle of the component X and P_{tx} is the probability of successful packet transmission.

$$Packets_{(fwd)}(t) = 1/T_{app} \times N_{sons} \times t + Packets_{(fwd)}(t_o) \quad (2)$$

$$E = E_{mac} \times D_{mac} + E_{beacons} \times D_{beacons} + E_{app} \times D_{app} \times 1/P_{tx} + E_{app} \times D_{app} \times (1+1/P_{tx}) \times N_{sons} + E_{sleep} \times [1 - D_{mac} - D_{beacons} - D_{app} - (1+1/P_{tx}) \times N_{sons} \times D_{app}] \quad (3)$$

The energy consumption model for a static network will be applied to a real-life, dynamic network when the network is stable, and it will be tested against real data to evaluate the error of the approximation. To simplify the model, the probability of successful packet transmission will be set to

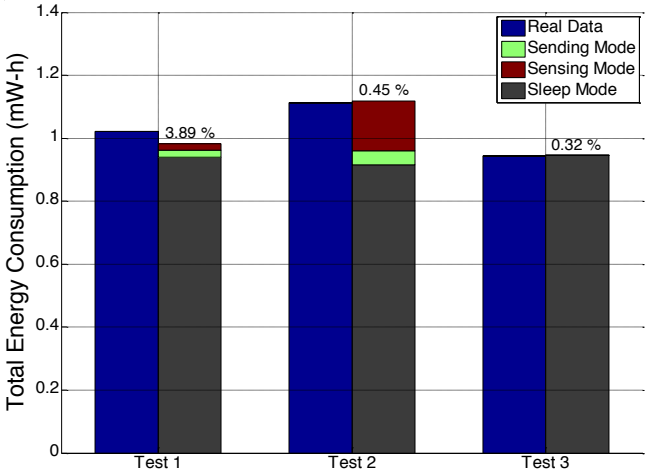


Figure 3. Comparison of the total energy consumption between the real data and the energy consumption model spliting into its components.

one. To consider any possible congestion in the network, a congestion factor "C" for forwarding packets will be introduced in (3). This is a probabilistic factor that resizes the energy consumption model depending on the network congestion. When the one of the nodes is congested its forwarding buffer fills up and the node starts dropping packets. This means that while sons keep sending packets, the parent node will not spend energy forwarding them and this has to be considered in the model. The congestion factor is the complement of an occupancy factor, which represents the rate of forwarding packets divided by the maximum rate that can be performed; the rate is the number of descendants for that node divided by the application activity period and the maximum rate is represented by the inverse of the MAC check period. Combining all together, the congestion factor is shown in (4).

$$C = 1 - [N_{sons} \times T_{mac} \ / \ T_{app}] \quad (4)$$

The final Energy Consumption Model obtained after these modifications is as shown in (5). This will be the model used and tested for validation.

$$E = E_{mac} \times D_{mac} + E_{beacons} \times D_{beacons} + E_{app} \times D_{app} + E_{app} \times D_{app} \times 2 \times C \times N_{sons} + E_{sleep} \times [1 - D_{mac} - D_{beacons} - D_{app} - 2 \times N_{sons} \times D_{app}] \quad (5)$$

B. *Validation of the Model*

In this section, the energy model consumption stated in (5) will be tested against real data and then a relative error will measure the accuracy of the model. For that, a dynamic wireless sensor network was set up with 6 nodes. After stabilization, the topology is as shown in Figure 4. The energy consumption on the forwarding node 1 has been monitored with a power analyser; node 0 is the base station and the network has a maximum of three hops. The Collection Tree Protocol (CTP) is responsible for routing and the application collects intrinsic network parameters such as "timestamp of the collection", "id of the node", "total number of sent packets", "number of packets sent to node X (array)", "total number of packets forwarded", "number of packets forwarded to node X (array)", "number of forwarding request" and "number of unacked packets". After stabilization, the energy consumption on node 1 was monitored for thirty minutes. I_{mac}, I_{sleep} $I_{beacons}$,and I_{app} were

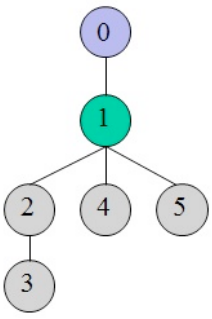


Figure 4. Network topology after stabilization.

empirically determined. Considering that the length of a beacon pulse must be at least equal to T_{mac} as beacons are broadcasted and they should be listened by all nodes in the network, the duty cycle for every component on the WSN application is as follows:

- D_{mac} = 4.75%, $D_{beacons}$ = 1.6%, D_{app} = 0.95% and D_{fwd} = 2.57%

Where D_{fwd} is 2 x C x N_{sons} x D_{app}, C is 0.34 and N_{sons} is 4.

Equation (2) was modeled and tested against real measurements after stabilization. Figure 5 shows the number of packets forwarded by node 1 and node 2 (forwarding nodes on the network) during the experiment. The error is very low and the model fits well with the real data, so it can be stated that (2) is valid and forwarded packets in a dynamic network can be approximately predicted using a model to predict forwarded packets in a static network when the environment is stable. Therefore, the duty cycle for forwarded packets in a dynamic network can be roughly estimated as well as the energy consumption produced by this component.

Figure 6 is a comparison of the real energy consumption and the energy predicted by the Energy Consumption Model. The predicted energy is disambiguated by its duty cycle, with each main activity represented on the model. Again, the major energy consumption is in sleep mode (it can be changed for low power devices), then the periodic medium check, forwarding packets (which is saturated by the congestion factor as the network is congested), sending beacons and the own application activity (which is very low due to simplicity of the tested application: it collects some parameters and sends them to the parent). The figure gives more detailed information about the energy consumption than if only the active mode and the sleep mode were considered.

Figure 7 is an indicative of the possible congestion on the network. It shows the number of unacknowledged packets on the different nodes. Node 3 has a large amount of unacknowledged packets that could be a consequence of dropped packets at Node 2; the same can be said about nodes 2, 4 and 5. Therefore, it may be possible that Nodes 1 and 2 are congested and the congestion factors need to be applied. Finally, to validate the model, the relative error between the model and the real data must be determined. Figure 8 compares total energy values measured and modeled. The

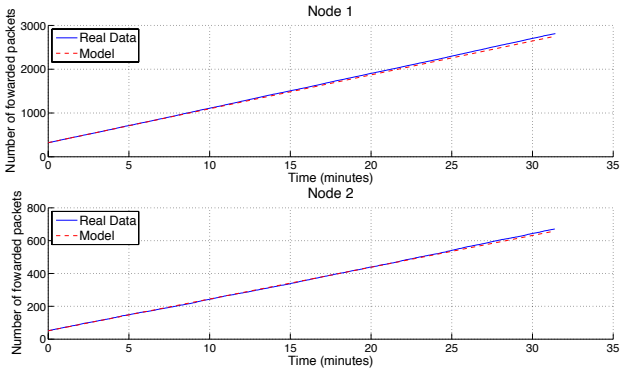


Figure 5. Number of forwarded packets per node and per time.

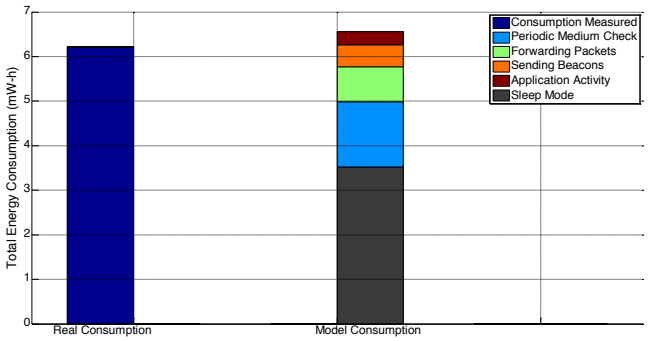


Figure 6. A comparison of the real consumption and the Energy Model Consumption by duty cycles, for node 1.

model equation presented on (5) has a relative error of 5.4 %. Also, the experiment was simulated using Matlab/Simulink based on (5). The congestion factor was modeled introducing the occupancy factor in a Bernoulli binary generator block. With this, the relative error drops to 3.1408 %. This simulation gives more detailed information about the error evolution over the time, Figure 8. Absolute error is linearly dependent on time, however, the relative error tends to be constant (around 3.2%).

V. DISCUSSION

A more precise energy control can be carried out using the energy consumption model presented in this paper. Thus, WSN applications can be customized and modeled to satisfy energy constraints decreasing duty cycles of each application component. Figure 6 gives a good indication of what consumes the most current and therefore it is possible to adjust the functionality of application components to reduce it. For instance, in this case, it can be stated that, in order to reduce energy consumption, the sleep consumption should be reduced but also the periodic medium check duty cycle should be reduced as the network is stable and the energy consumption for this component is considerably high. The application activity is low enough so there is no need of adjusting it.

However, there is always a trade off on reducing the energy consumption: decreasing the MAC duty cycle, the network will have more latency and might break down; likewise, reducing the beacon duty cycle, the network will take longer to adapt to changes. Therefore, it is needed to highlight here that the energy consumption model presented

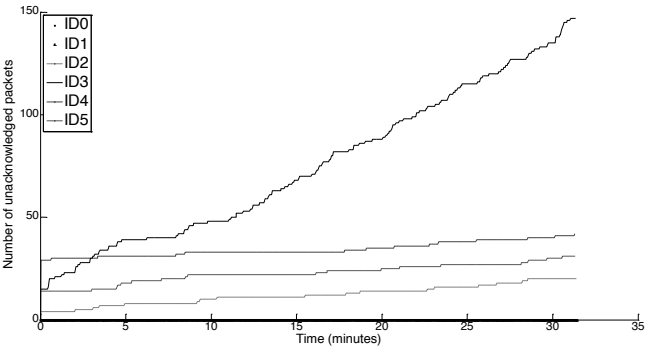


Figure 7. Number of unacknowledged transmissions per node.

TABLE I. COMPARATIVE TOTAL ENERGY OF THE EXPERIMENT FOR A WSN NODE.

	Total Energy comparative		
	Energy Consumption (mW-h)	*Absolute Error (mW-h)*	*Relative Error (%)*
Real Data	6.219	-	-
Model Equation	6.5553	0.33637	5.4088
Model Simulink	6.4143	0.19532	3.1408

in this paper does not aim to reduce or to give a fine-grained measure of the energy consumption but to discretize the energy consumption to determine what component consumes the most and how the consumption is distributed.

VI. FUTURE WORK

External factors such as symbolic interference, noise and multipath effect have a strong effect over the network communications and topology and, consequently over the energy consumption. In future work, a study of external factors will be carried out in order to include them in the energy consumption model. These factors will be modeled using Link Quality Indicator (LQI), Ambient Noise and Received Signal Strength Indication (RSSI). Also, deploying the network, it should be considered that, depending on the distance, links between two nodes could be in one of three statuses: connected region, transitional region and disconnected region. The packet reception rate (PRR) becomes quite random on the transitional region so, for future work, it is needed to avoid this region [13].

VII. CONCLUSIONS

An energy consumption model for a WSN node based on the disambiguation of the duty cycle has been achieved and presented on this paper. The division of the energy consumption in different components provides a better understanding of the internal process in nodes and allows network deployers to customize the overall consumption by modifying the functionality of the individual components as well as detect any malfunction that produces overconsumption.

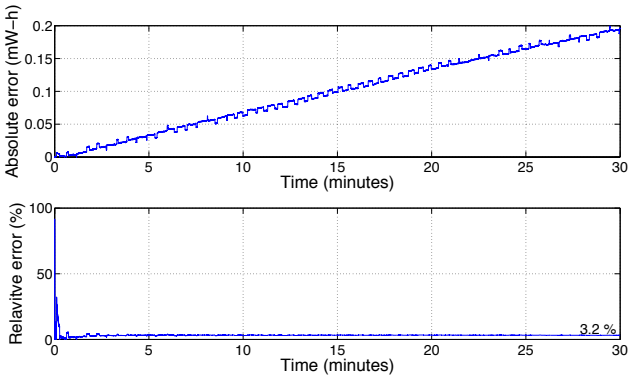


Figure 8. a)Absolute Error over the experiment. b)Relative Error over the experiment. Data measured for node 1.

 ISBN: 978-1-61208-296-7

The error obtained is low enough: using the model equation, the relative error is 5.4%; using a more stochastic process in Matlab/Simulink, the error drops until 3.14%. This model provides a more detailed view of the energy consumption than other models that use only active and sleep duty cycle.

ACKNOWLEDGMENT

This work has been supported by the Spanish Ministry of Economy and Competitiveness (LORIS project, ref. TIN2012-38080-C04-01), the FPI/UAH fellowship programs and the Science Foundation Ireland (ITOBO Strategic Research Cluster).

REFERENCES

[1] R. Lin, Z. Wang, and Y. Sun, "Energy efficient medium access control protocols for wireless sensor networks and its state-of-art," *IEEE International Symposium on Industrial Electronics*, May 2004, pp. 669–674 vol. 1.

[2] K. Langendoen, A. Meier, and E. T. H. Zurich, "Analyzing MAC protocols for low data-rate applications". *ACM Transactions on Sensor Networks (TOSN),* August 2010 vol. 7, no. 1.

[3] M. Younis, M. Youssef, and K. Arisha, "Energy-aware routing in cluster-based sensor networks," *Proceedings. 10th IEEE International Symposium on Modeling, Analysis and Simulation of Computer and Telecommunications Systems*, 2002, pp. 129–136.

[4] L. Krishnamachari, D. Estrin, and S. Wicker, "The impact of data aggregation in wireless sensor networks," *Proceedings 22nd International Conference on Distributed Computing Systems Workshops*, 2002, pp. 575–578.

[5] R. J. M. Vullers and R. Van Schaijk, "Energy harvesting for autonomous wireless," *Solid-State Circuits Magazine, IEEE,* 2010, pp. 29–38.

[6] A. Kansal, J. Hsu, S. Zahedi, and M. B. Srivastava, "Power management in energy harvesting sensor networks," *ACM Transactions on Embedded Computing Systems*, Sept. 2007, vol. 6, no. 4, p. 32–es.

[7] A. Kansal and M. B. Srivastava, "An environmental energy harvesting framework for sensor networks," *Proceedings of the 2003 International Symposium on Low Power Electronics and Design. ISLPED '03*, 2003, pp. 481–486.

[8] C. M. Vigorito, D. Ganesan, and A. G. Barto, "Adaptive control of duty cycling in energy-harvesting wireless sensor networks," *Sensor, Mesh and Ad Hoc Communications and Networks,* June 2007, pp. 21–30.

[9] K.-W. Fan, Z. Zheng, and P. Sinha, "Steady and fair rate allocation for rechargeable sensors in perpetual sensor networks," *Proceedings of the 6th ACM conference on Embedded network sensor systems - SenSys '08*, 2008, pp.239-252, 2008.

[10] E. Perla, A. Ó. Catháin, R. S. Carbajo, and M. Huggard, "PowerTOSSIM z : Realistic Energy Modelling for Wireless Sensor Network Environments," *Proceedings of the 3nd ACM workshop on Performance monitoring and measurement of heterogeneous wireless and wired networks,* 2008, pp. 35–42.

[11] Q. Wang, M. Hempstead, and W. Yang, "A realistic power consumption model for wireless sensor network devices," *2006 3rd Annual IEEE Communications Society on Sensor and Ad Hoc Communications and Networks*, 2006, pp. 286–295.

[12] K. Lorincz, B. Chen, J. Waterman, G. Werner-allen, and M. Welsh, "Resource aware programming in the pixie OS," *Proceedings of the 6th ACM conference on Embedded network sensor systems,* 2008, pp. 211–224.

[13] N. Baccour, A. K. Aa, L. Mottola, H. Youssef, C. A. Boano, and M. Ario, "Radio Link Quality Estimation in Wireless Sensor Networks : a Survey," *ACM Transactions on Sensor Networks (TOSN),* Sept. 2012, vol. 8, no. 34.

 ISBN: 978-1-61208-296-7

A New Clustering Algorithm in WSN Based on Spectral Clustering and Residual Energy

Ali Jorio*, Sanaa El Fkihi † Brahim Elbhiri‡and Driss Aboutajdine*
*LRIT, Research Unit Associated to the CNRST (URAC 29), FSR, Mohammed V-Agdal University, Rabat, Morocco
jorio.ali@gmail.com, aboutaj@fsr.ac.ma
†ENSIAS Mohammed V University Souissi Rabat, Morocco
elfkihi@ensias.ma
‡EMSI Rabat, Morocco
elbhirij@emsi.ac.ma

***Abstract*— Wireless Sensor Networks (WSNs) are composed of large number of sensor nodes that are randomly distributed in a region of interest. The nodes are responsible of the supervision of a physical phenomenon and periodic transmission of results to the sink. Energy saving results in extending the life of the network, which presents a great challenge of WSNs. To deal with this, a hierarchical clustering scheme, called K-Way Spectral Clustering Algorithm in Wireless Sensor Network (KSCA-WSN), is proposed in this paper. This algorithm is based on spectral classification; it also considers the residual energy, as well as some properties of the network nodes. Thus, our approach aims to seek for an ideal distribution of sensor nodes (clusters) and proposes new feautures to elect the appropriate cluster-heads. In term of extending the network lifetime and minimizing the energy consumption, the simulation results show an important improvement of the network performances with KSCA-WSN compared to other existing clustering methods.**

Keywords-*Clustering; Graph theory; Spectral classification; Energy consumption.*

I. INTRODUCTION

A Wireless Sensor Network (WSN) is a wireless network consisting of spatially distributed autonomous devices using sensors to cooperatively monitor physical or environmental conditions at different locations, such as temperature, sound, and vibration. It contains a large number of nodes, which sense data from an inaccessible area and send their reports towards a processing center called Base Station (BS). WSNs have many advantages, such as the ease deployment and the capacity of self-organization. However, their main challenges are the limited resources in terms of radio communication capabilities and energies. This article addresses this last point.

Even if the sensors are usually powered by batteries, it is not practical to recharge or replace them, because they are often deployed in hostile environments. Furthermore, in a WSN, a large part of energy is consumed when communications are established [1]. Hence, frequent and long distance transmissions should be minimized to extend the lifetime of the network. One way to reach this result consists in dividing the network into multiple clusters; each cluster has a cluster-head node (C-H) [2]. The C-H collects data from nodes of its same cluster, aggregates them and transmits them to the BS. Nevertheless, the main problem of many proposed clustering protocols is the random selection of the C-Hs. Indeed, all C-Hs can be located in a small region of the network, and some ordinary nodes will be isolated. To tackle the clustering problem and to fairly localize C-Hs, we propose to consider methods of spectral clustering and the sensor's residual energies.

Spectral clustering has become one of the most popular modern clustering algorithms. It is based on graph theory, can be solved efficiently by standard linear algebra software. It also and very often outperforms traditional clustering algorithms such as the k-means algorithm [3]. It includes a variety of methods based on the notion of similarity matrix and use the eigenvectors differently. The k-ways approach is one of these methods. It consists to divide points into K disjoint classes. This method is based on the K eigenvectors related to K largest eigenvalues of Laplacian matrix [4]. Following this approach, we present a K-Way Spectral Clustering Algorithm in Wireless Sensor Network (KSCA-WSN), which has four main goals: (i) extending the network life-time by distributing energy consumption, (ii) running the clustering process within a constant number of iterations/steps, (iii) minimizing control overhead (to be linear in the number of nodes), and (iv) producing well-distributed cluster heads and compact clusters. KSCA-WSN does not make any assumptions about the distribution or density of nodes, neither about node capabilities.

The remaining of this paper is organized as follows: In Section II, we give a brief overview of some related research work. Section III describes the network model and states the addressed problem in this work. Details and properties of the proposed algorithm are given in Section IV while Section V presents the simulations and the results. Finally, conclusion and some perspectives are drawn in Section VI.

II. RELATED WORK

Many routing protocols have been designed based on clustering [5], where C-Hs are elected periodically and alternately. Moreover, the most commonly used approach for clustering is the LEACH algorithm [6]. LEACH is an energy-efficient communication protocol, which employs a hierarchical clustering. Besides, many clustering protocols based on the principle of this algorithm have been developed, such as LEACH-C [7], Energy-Kmeans [8], DECSA [9], EECS [10] and SECA [11].

 ISBN: 978-1-61208-296-7

In LEACH, nodes organize themselves into clusters using a distributed algorithm. Its main idea is to randomly and periodically select the C-Hs. The probability of becoming a C-H for each period is chosen to ensure that every node becomes a C-H, at least once within N/K rounds, where N is the number of node and K is the desired number of clusters. After the election of C-Hs, each ordinary node in the WSN determines its cluster by choosing the C-H that requires the minimum communication energy, based on the received signal strength of the advertisement from each C-H. Each ordinary node will choose to join the C-H which has the highest signal quality. Once the clusters are formed, each C-H node creates Time division multiple access (TDMA) schedule. The C-H collects and aggregates information from sensors in its own cluster and passes on information to the BS. Although LEACH distributed clustering algorithm has advantages, using a central control algorithm to form clusters may produce better clusters by dispersing the cluster head nodes throughout the network [7]. In LEACH-centralized (LEACH-C), the BS determines the C-Hs by computing the average node energy. The node with energy below this average cannot be cluster heads for the current considered round. This ensures that the energy consumption will be effectively distributed among all the nodes.

Yong and Pei [9] presented a distance-energy cluster structure algorithm (DECSA) based on the classic clustering algorithm LEACH. This proposed approach considers both the distance between the nodes, the position of the BS and residual energy of nodes. Its main idea is to partition the network into three levels of hierarchy to reduce the energy consumption of C-Hs. this results from the non-uniform distribution of nodes in the network and thus avoid direct communications between the BS and C-H that has minimal energy and is far away from the BS.

Elbhiri et al. [12] propose a new approach called the Spectral Classification based on Near Optimal Clustering in Wireless Sensor Networks. This approach is based on spectral bisection for partitioning a sensor network into two clusters. This could be applied recursively to get the desired number of clusters. However, recursive spectral bisection finally produces 2^n clusters where n is the number of iterations. This method cannot partition the network into any desired number of clusters.

III. RADIO MODEL AND PROBLEM STATEMENT

This paper, discusses the periodical data gathering application, for which LEACH is proposed. Thus, the proposed clustering algorithms usually produce clusters of the same size. The C-H consumes the same amount of energy during intra-cluster communications. Moroever, cluster head election method based essentially on residual energy can obtain better energy efficiency than the method in which cluster heads are elected in turns or by probabilities. For this, we use the spectral classification that is widely used to solve graph partitioning problems. The spectral methods get their name from the spectral theories of linear algebra. This theories allows affirming the diagonalization of real symmetric matrices. It also justifies the decomposition of real symmetric matrices into eigenvalues in an orthonormal set of eigenvectors. Besides, the graph partitioning problem can be reduced to the resolution of a numerical system $Mx = \lambda x$. Solving this numerical system consists in seeking an orthogonal base of eigenvectors of the matrix M [4].

In our approach, the BS constructs the graph corresponding to the WSN based on the spectral clustering principle. Indeed, let x be an observation vector composed of the sensor network nodes (let N be the number of nodes). This vector can be represented by an undirected graph $G(V, E)$; where V is the set of vertices (sensor nodes) identified by an index $i \in \{1, \cdots, N\}$ and E is the set of edges that link each two vertices (communication link). Let $A \in \Re^{N \times N}$ be the similarity matrix of the graph G.Each value of A is associated to each pair of the graph nodes (i, j). The total weight of incident edges to node i is given by $d_{ii} = \sum_{j=1}^{j=N} a_{ij}$. The degree matrix $D \in \Re^{N \times N}$ of G is a diagonal matrix defined by $D = [d_{ij}]$. Finally, the Laplacian matrix of the graph is calculated by :

$$L = D^{-\frac{1}{2}} * A * D^{-\frac{1}{2}} \quad (1)$$

To simplify the network model, we make some assumptions that are:

- All nodes are homogeneous and have the same capacities.
- Each node is assigned a unique identifier (id) that includes the cluster identifier to which the node belongs.
- The network topology remains unchanged over time.
- The BS is placed far away from the network.

In addition, we use the model of the radio hardware energy dissipation presented in [6] and [7], and illustrated by Figure.1.

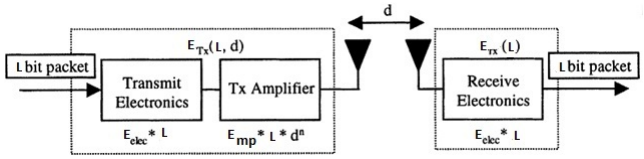


Fig. 1. Radio energy dissipation model.

Thus, to transmit an $L - bits$ message between two nodes separated by a distance d, the spent energy by the radio transmitter is given by the next equation:

$$E_{Tx}(L, d) = \begin{cases} L * E_{elec} + L * E_{fs} * d^2 & if\ d < d_0 \\ L * E_{elec} + L * E_{mp} * d^4 & if\ d \geq d_0 \end{cases} \quad (2)$$

and to receive this message, the spent energy by the radio receiver is:

$$E_{Rx}(L, d) = L * E_{elec} \quad (3)$$

where, E_{elec} is the energy consumption per bit in the transmitter and receiver circuitry, E_{fs} is the free space model's

 ISBN: 978-1-61208-296-7

amplifier energy consumption, and E_{ms} is the multiple attenuation model's amplifier energy consumption. The threshold's value (d_0) is determined by the equation (4):

$$d_0 = \sqrt{\frac{E_{fs}}{E_{ms}}} \tag{4}$$

In addition, once a sensor node runs out its energy, it is considered as died.

IV. KSCA-WSN Details

In this section, we give details of the proposed KSCA-WSN algorithm. The latter consists in partitioning the WSN into a set of clusters based on spectral classification. Besides, it determines the CHs based on the residual energy of each node of the network. Some other properties will be explained in the subsections.

The spectral classification algorithm proposed here consists of three steps, as illustrated in the Figure 2.

A. The Pre-Processing step

First, each wireless sensor network node determines its position by using a Global Positioning System (GPS) or other tracking device, and transmits it to the BS in a short message. Based on the spectral clustering principle, the BS constructs the graph G and the similarity matrix A which represents the network. Each value of A is associated to each pair of the graph nodes (i, j). This value is of gaussian type and the matrix A is given by eq.(5).

$$A = [a_{ij}] = \begin{cases} exp\left(\frac{-1}{2\sigma^2} * d^2(i,j)\right) & if\ i \neq j \\ 0 & otherwise \end{cases} \tag{5}$$

with $d(i, j)$ is the euclidian distance between nodes i and j.

Thereafter, the BS must deduce the Degree Matrix and the Laplacian one. Then, it computes the eigenvalues and the eigenvectors of the last matrix.

B. Clustering step

The objectives of the current step are to define the optimal number of clusters and to form them.

As indicated in [7], the optimum number of clusters can be determined by the next equation:

$$K = \frac{\sqrt{N}}{\sqrt{2\pi}} * \sqrt{\frac{E_{fs}}{E_{ms}}} * \frac{M}{d_{toBS}^2} \tag{6}$$

where, N is the number of sensor nodes, M is the area of sensor nodes deployment, and d_{toBS} is the average distance of the nodes from the BS.

We construct a new matrix U from the K eigenvectors related to the K largest eigenvalues of the Laplacian matrix. In order to determine the K clusters of the WSN, we apply the classification algorithm k-means to the matrix U. We deal with each row of U as a point in $\Re^K$. We cluster the WSN into K clusters via k-means. The sensor node i is assigned to cluster C_j if and only if row i of the matrix U was assigned to cluster C_j.

We notice that in the proposed algorithm we determine the clusters before specifying the cluster heads and the optimal number of cluster partitions is as well defined automatically. So, our algorithm is different from the others (such as LEACH, LEACH-C, DECSA, etc.) that determine the C-Hs before the clusters. In addition, in order to define the number of cluster heads, the latter protocols run the same technique in each iteration and by the way consume more energy.

C. Cluster head election step

Once the clusters are determined, the next step of the KSCA-WSN consists to select the C-Hs. The C-Hs are responsible of the coordination among the nodes within their clusters, as well as for the aggregation of the received information and the communication with the BS. In our algorithm, the C-H is determined by taking into account the id of the node in the cluster and its residual energy. The id is defined without considering the position of the node in the cluster; thus, the C-H in each round of communication will be at a random position on the cluster. It is so important that nodes die at random locations of the network.

Indeed, for each round r of the simulation, we use the number $C_k = (r\ mod\ |S_k|)$ to elect a C-H for the appropriate cluster; $|S_k|$ is the total number of nodes in the cluster k. The node with $id = C_k$ and residual energy E_r greater than threshold E_{rmin} will be the C-H of the cluster k in the round r. E_{rmin} is the minimum residual energy required for a given node to be a C-H. It is the summation of the energy needed to receive and process data coming from the appropriate cluster nodes (data aggregation), and to transmit data towards the BS. This energy E_{rmin} is given by the next equation:

$$E_{rmin} = |S_k| * (E_{Rx}(L,d) + E_{Aggregation}) + E_{Tx}(L,d) \tag{7}$$

with

- $E_{Tx}(L,d)$is the energy consumed when the C-H transmits $L - bits$ data to the BS by a distance d.
- $E_{Aggregation}$ is the energy needed by the C-H to process data :

$$E_{Aggregation} = L * E_{DA} \tag{8}$$

  E_{DA} is the energy for data aggregation per bit.
- $E_{Rx}(L,d)$is the energy consumed when the C-H receives data from one node.

Hence

$$E_{rmin} = L * \big((|S_k| + 1) * E_{elec} + |S_k| * E_{DA} + \epsilon_{mp} * d_i^4\big) \tag{9}$$

Nevertheless, if the residual energy E_r is less than this threshold E_{rmin} this node must broadcast a short message informing the node with $id = C_k + 1$to its residual energy and so on. Thus, the energy consumption will be distributed with more equitability between all nodes.

D. Data Transmission

Based on the id and the numbers of nodes in the cluster, a schedule TDMA will be created automatically to assign to each node a time when it can transmit its data to its own C-H. If we suppose that the node with the $id = i$ is elected

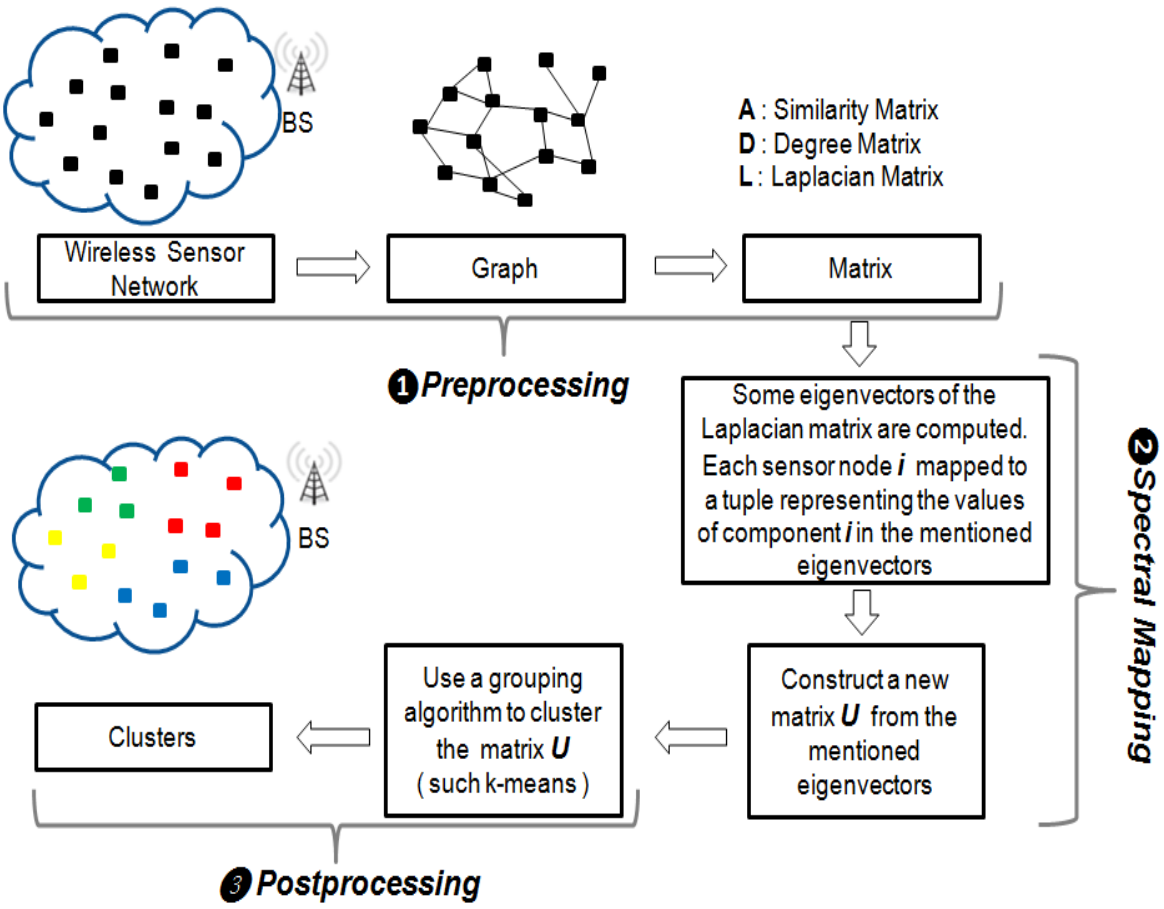


Fig. 2. Spectral clustering scheme of the KSCA-WSN approach.

as a C-H, the node with $id = ((i + 1 + |S_k|) \ mod |S_k|)$ is assigned the first period to transmit its data.

One of the most important challenges in our approach consists in reducing the total consumed energy of each round. Hence, we avoid energy consumption due to synchronization of the cluster nodes when the C-H is elected to assign the TDMA. Indeed, this technique ensures not having collisions between data messages. It also allows the radio components of each non-C-H node to be disabled at any time, except during the time it is allocated for transmitting these data. this reduces the energy consumed by the nodes. However, the C-H must keep its receiver in order to receive data from other nodes. In addition, to save more energy in a wireless sensor network, we assume that if the distance between a node and the BS is lower than the distance between this node and its C-H. Then the node transmits its data directly to the BS. Moreover, each node can move in the standby mode to reduce energy consumption.

V. Simulations and Results

In this section, we present the results of the evaluation experiments of the proposed KSCA-WSN algorithm. We used the MATLAB software to simulate and analyze this algorithm. Besides, we consider the radio model presented in Figure 1 and the different parameters used in our simulations are shown in TABLE I.

TABLE I
Experimental Simulation parameters.

Parameter	value
E_{elec}	$50nJ/bit$
E_{fs}	$10pJ/bit/m^2$
E_{mp}	$0.0013pJ/bit/m^4$
Initial energy E_0	$0.5J$
E_{DA}	$5nJ/bit/message$
Area of Network	$200m * 200m$
Sink coordination	$(100m, 250m)$
d_0	$88m$
Message Size	$4000bytes$
Number of Nodes	500

A node that runs out of energy is considered dead and cannot transmit or receive data. For these simulations, the energy of a node decreases each time it sends, receives or aggregates data according to the model radio parameters. Also, we ignore the effect caused by the signal collision and the interference in the wireless channel. Furthermore, each simulation result

 ISBN: 978-1-61208-296-7

shown below is the average of 100 independent experiments, where each experiment uses a different randomly-generated uniform topology of sensor nodes.

As illustrated in Figure 3, an example of wireless sensor networks with $N = 500$ nodes randomly distributed in a $200m*200m$ area. The BS is located far away from the sensing area $(x_{SB} = 100m\ ;\ y_{SB} = 250m)$.

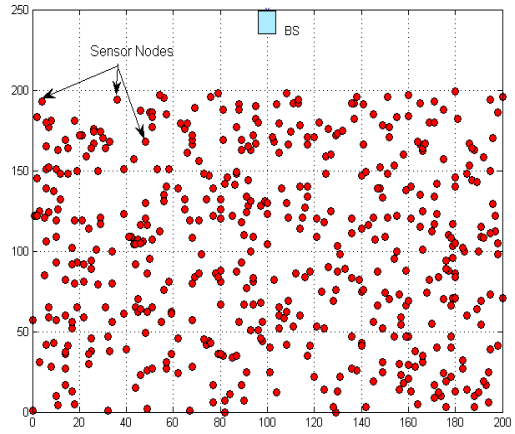


Fig. 3. Distribution of wireless sensor network.

In order to evaluate the performances of the new proposed protocol, we propose to compare it to the Spectral Classification based on Near Optimal Clustering (SCNOC) [12], the DECSA [9] and the LEACH-C [7] algorithms. Also, we use two metrics to analyze and compare the performances of the protocols:

- Network lifetime: It can be defined as the time elapsed until the first node (or the last node) in the network depletes its energy (dies). This paper uses First Node Dies (FND) metrics which were presented as an approach to define lifetime of a WSN in literature.
- Energy consumption: Uniform energy consumption is very important for network load balancing. More uniform energy consumption, less possibility for node premature death. And less energy consumption per round, more network performance.

As illustrated in Figure 4, the main problem of LEACH-C and DECSA protocols are the random selection of the CHs. Indeed, all CHs can be located in a small region of the network. Hence, some ordinary nodes would be out of reach. Also, SCNOC produces 2^n clusters where n is the number of iterations; it cannot partition the network into any desired number of clusters.

The Figure 5 presents an example of clustering for the KSCA-WSN algorithm. We note that the network is subdivided into six clusters, and the nodes are correctly distributed over the sensing area. Also, there is no intersection between the different clusters.

Figure 6 shows a significant improvement for the KSCA-WSN approach in terms of numbers of periods relating to

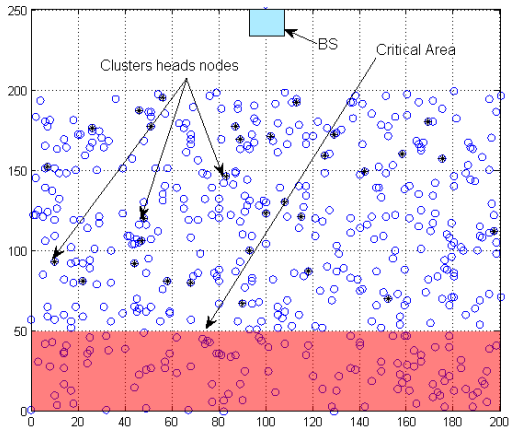


Fig. 4. Illustration of the critical area in the LEACH-C clustering algorithm with N=500 nodes

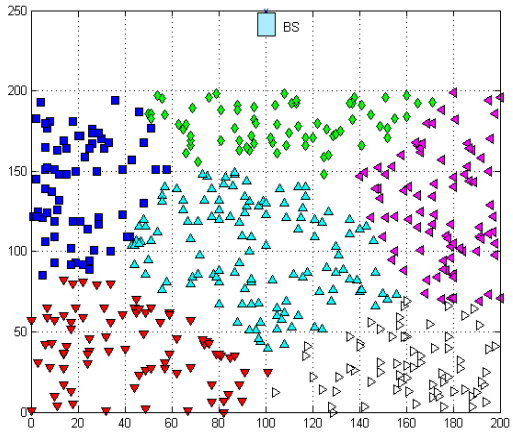


Fig. 5. Clustering results of the KSCA-WSN algorithm with N=500 nodes.

FND. In the proposed method, FND occurs at the period 520 while it occurs at the 378 period for spectral classification based on Near Optimal Clustering, at 21 periods for LEACH-C, and at 74 for DECSA protocols.

Figure 7 gives the total network remaining energy in every transmission round. The network remaining energy decreases more rapidly in the spectral classification based on Near Optimal Clustering, the LEACH-C, and the DECSA protocols than in the KSCA-WSN algorithm.

The Figure 8 shows the effects of the node density on the compared clustering techniques as well as on the network's stable regions (First Node Dead 'FND').

As shown in Figure 8, for different values of N equal to 100, 200, 300, and 400, our algorithm presents an improvement of performances compared to the other algorithms. It follows that even if the node density increase the new proposed approach still gives best results compared to the other ones.

Figure 9 shows the performances of the different compared

 ISBN: 978-1-61208-296-7

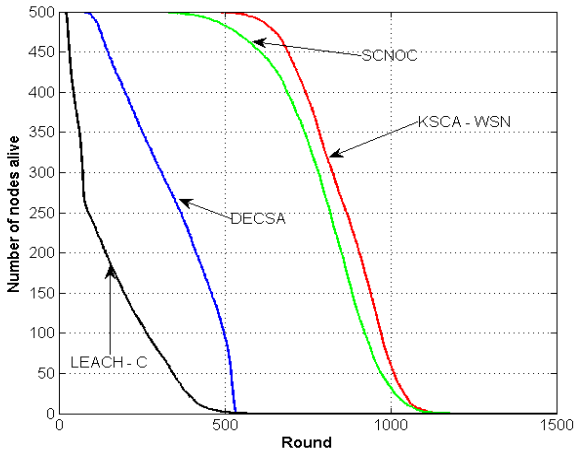


Fig. 6. Number of nodes alive over time of the compared protocols

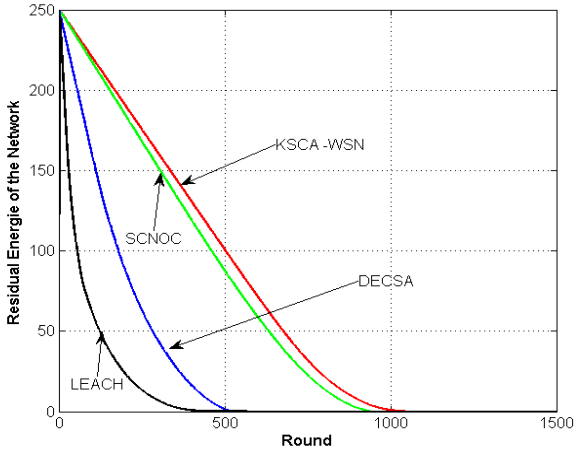


Fig. 7. Evolution of the remaining energy in the network when the transmission rounds.

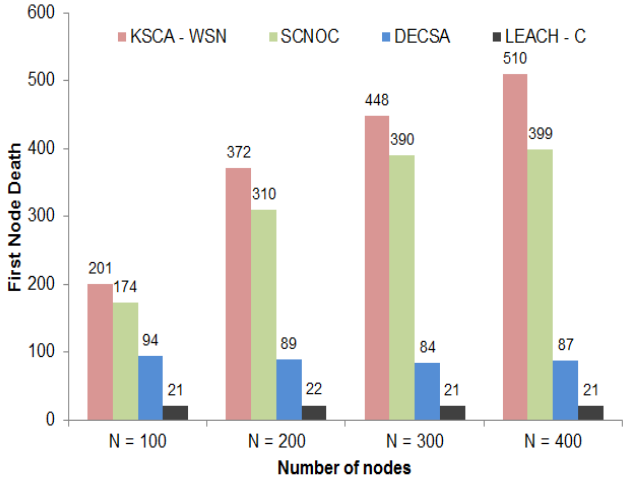


Fig. 8. Impact of the node density N on the performances of the compared algorithms

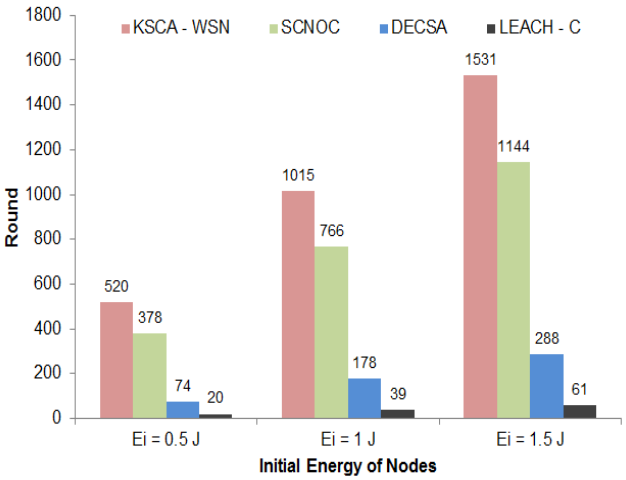


Fig. 9. Impact of the initial energy quantity on the performances of the four compared algorithms

protocols by using different initial energies. It gives the FND round depending on the quantity of the node initial energy. Once more, it is shown that for different values of the energy, the new proposed approach presents a significant improvement compared to the others.

We conclude that the KSCA - WSN algorithm gives a significant performance improvement; in terms of energy and lifetime gains, compared to the Spectral Classification based on Near Optimal Clustering, the LEACH-C, and the DECSA protocols. The best results of the KSCA-WSN approach can be explained by the three points: (i) the proposal starts by selecting the clusters (similar nodes) before the election of the C-Hs. (ii) the approach considers the WSN nodes residual energies when electing C-Hs. And (iii) a C-H is elected at any position of the wireless sensor network.

VI. Conclusion and Future Work

Energy saving is an important challenging issue in a wireless sensor network. To ensure more effective energy distribution and to extend the life time of a network, new and efficient energy saving scheme is developed in this paper. We have proposed a new method of clustering, based on spectral classification. The latter has been widely used to solve graph partitioning problem. In particular, we use the method of k-ways. Thus, we start by determining similar nodes (clusters) before selecting C-Hs. Furthermore, we take into account the node residual energy when selecting the appropriate C-H nodes. Indeed, simulation results show that our approach ensures the low energy consumption and improves the network lifetime. Promising results are obtained by several simulations, and exceed those obtained by others algorithms.

Further works need to determine other properties to improve both the clustering and the cluster-head election processes.

References

[1] I. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "A survey on sensor networks," Communications magazine, IEEE, vol. 40, no. 8,

 ISBN: 978-1-61208-296-7

2002, pp. 102−114.
[2] J. Al-Karaki and A. Kamal, "Routing techniques in wireless sensor networks: A survey," Wireless Communications, IEEE, vol. 11, no. 6,2004, pp. 6−28.
[3] U. Von Luxburg, "A tutorial on spectral clustering," Statistics and computing, vol. 17, no. 4, 2007, pp. 395−416.
[4] A. Ng, M. Jordan, Y. Weiss, "On spectral clustering: Analysis and an algorithm," Advances in neural information processing systems, vol. 2, 2002, pp. 849−856.
[5] O. Boyinbode, H. Le, and M. Takizawa, "A survey on clustering algorithms for wireless sensor networks," International Journal of SpaceBased and Situated Computing, vol. 1, no. 2, 2011, pp. 130−136.
[6] W. Heinzelman, A. Chandrakasan, and H. Balakrishnan, "Energyefficient communication protocol for wireless microsensor networks," in System Sciences, 2000. Proceedings of the 33rd Annual Hawaii International Conference on, vol. 2. IEEE, 2000, p. 10.
[7] W. B. Heinzelman, A. P. Chandrakasan, and H. Balakrishnan, "An application-specific protocol architecture for wireless microsensor networks," IEEE Transactions on wireless communications, vol. 1, no. 4, 2002, pp. 660−670.
[8] S. Jerusha, K. Kulothungan, and A. Kannan, "Location aware cluster based routing in wireless sensor networks," International Journal of Computer and Communication Technology, vol. 3, no. 5, 2012, pp. 0975−7449.
[9] Z. Yong and Q. Pei, "A energy-efficient clustering routing algorithm based on distance and residual energy for wireless sensor networks," Procedia Engineering, vol. 29, 2012, pp. 1882−1888.
[10] M. Ye, C. Li, G. Chen, and J. Wu, "Eecs: an energy efficient clustering scheme in wireless sensor networks," in Performance, Computing, and Communications Conference, IPCCC 2005, IEEE, 2005, pp. 535−540.
[11] J.-Y. Chang and P.-H. Ju, "An efficient cluster-based power saving scheme for wireless sensor networks," EURASIP Journal on Wireless Communications and Networking, vol. 2012, no. 1, 2012, pp. 172.
[12] B. Elbhiri, S. El Fkihi, R. Saadane, and D. Aboutajdine, "Clustering in wireless sensor networks based on near optimal bi-partitions," in Next Generation Internet (NGI), 2010 6th EURO-NF Conference on. IEEE, 2010, pp. 1−6.

 ISBN: 978-1-61208-296-7

Energy-Efficient Posture Classification with Filtered Sensed Data from A Single 3-Axis Accelerometer Deployed in WSN

Laurentiu Hinoveanu, Jacek Lewandowski, Xiang Fei, Hisbel Arochena, Partheepan Kandaswamy, Zhipeng Dai
Faculty of Engineering and Computing
Coventry University, Coventry, UK
{hinoveal, j.lewandowski, x.fei, h.arochena, kandaswp, daiz}@coventry.ac.uk

***Abstract*— Wireless Sensor Networks (WSNs) provide rich and detailed measurements of the physical phenomenon that they monitor. With the sensed data, a variety of pervasive applications can be developed. One category of those applications are classifications based on supervised machine learning, with one example being postures recognition with data from body sensor networks (BSNs). Conventionally for accuracy reason, raw data from the BSN sensors, such as accelerometers or other inertial devices, is transmitted to the central unit for postures identification. It has been well known, however, that in most of the battery powered WSNs, communication consumes most of the energy. This paper explores the possibility of obtaining the same level of classification accuracy with the filtered sensed data to prolong the lifetime of the WSNs. A special case of posture recognition based on Artificial Neural Networks (ANN), Naive Bayes and K-Nearest-Neighbours (KNN) has been studied, and a mechanism for the posture classification based on filtered sensed data has been constructed. Real data from a Shimmer node has been collected and the test results show that the same level of accuracy can be obtained using only two thirds of the raw data. The implementation considerations with some prototypes have also been provided.**

Keywords - Wireless Sensor Network; classification; Artificial Neural Network; Naive Bayes classifier; K-Nearest-Neighbours

I. INTRODUCTION

Wireless Sensor Networks (WSNs) provide rich and detailed data about the physical phenomenon measured by the sensors. With the sensed data, a variety of pervasive applications can be developed in areas such as environmental monitoring, industrial and manufacturing automation, health-care, and military [1]. Among others, one category of the applications are classifications based on supervised machine learning. One example of such a category is posture recognition with data from body sensor networks (BSNs) [2]. Conventionally for accuracy reasons, raw data from the sensors, such as accelerometers or other inertial devices, is transmitted to the central unit for training and then postures recognition. It has been well known, however, that WSNs communication is one of the main factors that contribute to the energy consumption. The cost, measured in energy units, of transmitting a single bit of information is approximately the same as that needed for processing a thousand operations in a typical sensor node [3] [4]. Although recent research efforts focus on energy harvesting which aim to convert the ambient energy from the environment into voltage, current technology is only enough to power the sensor node continuously for 1% to 20% of its operating time [5]. Moreover most of this algorithms is still in experimental and prototype stage [6] [7]. Therefore, before energy harvesting technology become mature, minimizing the amount of communication is still a most effective way of prolonging the lifetime of WSNs.

This paper explores the possibility of obtaining the same level of classification accuracy with the filtered sensed data compared to the raw data. Presented here BSN system, consisting of a single sensor node with one 3-axis accelerometer has been deployed for recognizing four basic postures. These are: standing, walking, running and kneeling. Presented study compares three different machine learning algorithms: ANNs, Naive Bayes, and KNNs. The obtained classification accuracy with raw sensed data for all three algorithms is over 80% what compares favourably with other published results that use single 3-axis accelerometer only. With the same BSN system, a mechanism of classifying using filtered sensed data has been constructed. Real data from test subjects have been collected and the test results show that the same level of accuracy can be obtained using only two thirds of the raw data transmitted from the sensor node.

The rest of the paper is organized as follows. Section II presents background on classification algorithms with filtered and raw data. Section III outlines our BSN system for postures recognition. In Section IV, the mechanisms for classification using both the raw data and the proposed

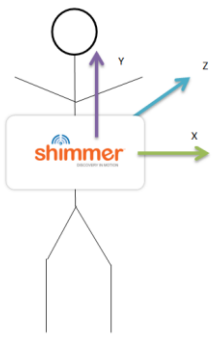


Figure 1. BSN system for postures recognition

 ISBN: 978-1-61208-296-7

filtered data are described. In Section V test results are evaluated and analysed. Section VI presents ongoing and future implementation works while section VII concludes the paper.

II. BACKGROUND

The classification techniques have been widely adopted in developing WSNs applications. Trabelsi et al. [8] formulated the activity recognition problem from 3-d acceleration data measured with body-worn accelerometers as a problem of multidimensional time series segmentation and classification. They compared supervised and unsupervised classification approaches for human activity recognition using body-mounted sensors

Long et al. [9] used the related Naive Bayes method for posture classification using a single tri-axial accelerometer placed on the waist, obtaining 80% accuracy.

Djuric-Jovicic et al. [10] designed an ANN based algorithm for automatic recognition and classification of walking patterns with a wireless inertial measurement system.

Archibald and Fann [11] studied the effect of the number of features used on the accuracy of a classification by an SVM (Support Vector Machine). The study is based on a series of classification analyses with two hyperspectral sensor data.

Godfrey et al. [12] demonstrated an application of the scalar (dot) product technique and vertical velocity estimates to detect transitions and their types with only a single tri-axial accelerometer. Activities and postural transitions were accurately detected by this simplified low-power kinematic sensor and corresponding activity detection algorithms. As reported they obtained a sensitivity and specificity of 86–92% for young healthy subjects in a controlled setting and 83–89% for elderly healthy subjects in a home environment.

All the classification mechanisms mentioned above take raw or nearly raw sensed data as their input.

Data filtering in WSNs applications normally refer to low-pass filtering [2] or extended Kalman filters (EKF) [13], etc., that run on the sensor nodes and operate on the current and historical sensed data.

In [1] and [14] the authors used information fusion techniques, more precisely data aggregation techniques, to reduce the amount of data traffic. Moreover, Boulis et al. [14] presented a distributed estimation algorithm that exploits the energy–accuracy trade-off and can be applied to a subclass of periodic aggregation problems. However, in all cases the aggregated data is to be sent to the remote base stations for machine learning based classification.

III. BSN SYSTEM FOR POSTURES RECOGNITION

Postures recognition can be used in many applications for instance to monitor the elderly in order to identify falls or other potential hazards to a person's well-being as a result of restricted mobility or other physical impairments. A simple prototype BSN system used in this study, consists of a Shimmer-2 sensor node equipped with a three-axis accelerometer [15]. The sensor node communicates wirelessly with the remote base station via Bluetooth radio. The device is attached to the subject's abdominal using an elastic band as shown in Figure 1. In this study, four basic body postures were chosen. These are: walking, running, standing and kneeling positions. The first two are dynamic postures, while the latter two are static ones. This set covers the main features of the postures that are of interest for elderly postures monitoring applications.

IV. METHODS

There are many classification mechanisms such as ANN based, Naive Bayes based, KNN based, among others which find their application to posture recognition.

The idea of an ANN is to straightforwardly quasi-simulate the fundamental activity of a network of neurons, which is based on potential build-ups' mechanism for biological neural communication. Therefore, in an ANN there exists a function that governs the signalling of an adjacent neuron in the next phase layer and a random "favouritism" factor in each neuron. Usually, that "signalling" function is chosen in such a way that its "first derivative" is easily computed, typically having a clear association to the original signalling function [16]. Some of the preferred functions are the exponential function or trigonometric functions, especially the sine and cosine that in the complex plane both contribute to the construction of the complex exponential and have easily recognisable derivatives. The weights of the interconnections are computed based on a "learning factor" provided by the user. These are normally within values between 0 and 1. It governs a result given by the multiplication of the previous function, the "learning rate", and a "fault term". These mechanism manages the variation in the values obtained from the test set and the training set as well as the previous results calculated during the training period. The weighting factors should be optimized so that the difference between the outcomes obtained from the unlabelled test set and the facts resulted from the training set is "minimal" [16].

Based on the understanding of set theory and the classical definition of the probability and of the conditional property of it, we could infer Bayes theorem as the ratio of a conditional probability of an event M_1 given an event M_2 and the conditional probability of M_2 assuming M_1 is equal to the ratio between probability for event M_1 and the probability for the second event, presuming that both denominators are not zeroes, otherwise those ratios would be mathematically undefined [17]. This explanation and a basic assumptions are the founding elements of the Naive Bayes classifier. The supposition is that the events that govern each of the samples in the training set are unrelated with each other, which simplifies the calculation of the mutual probability to just the multiplication of the probability for each instance in the training set [18]. KNN algorithm is also constructed on an assumption which corresponds to the principle that the largest quantity of samples, which are

 ISBN: 978-1-61208-296-7

all labelled under the same class [19] and are closest to the new to-be-classified case, gives the corresponding labelling [20]. For further information regarding ANN classifier, Naive Bayes classifier, and KNN classifier, please refer to [16] [18] and [19].

A. Postures Recognition with Raw Sensed Data

As the postures classification in this study is based on supervised machine learning, here we take ANN based classification as an example, and describe the general procedure as follows:

1. Determining the features used for postures classification: in our case, three dimensional accelerations are chosen as the input of the classifier;
2. Assigning values to the four postures as the output of the classifier;
3. Collecting the sensor data that measures the features of the four postures. Two sets of data should be collected, one for training (supervised learning) and the other for testing.
4. Designing the classifier: for ANN classifiers, the structure (or layers) of the neural networks, the connections between the neurons, and the activation function have to be determined. In our case, a three-layer feed-forward back-propagation ANN is adopted with full synapses in-between. The activation function is the sigmoid function.
5. Supervised machine learning
6. Testing the performance of the designed classifier on a group of new subjects;
7. Evaluating the test results: if the performance of the designed classifier is not acceptable, either re-design the classifier or pick up other features for the classification.

Conventionally, both training and testing phases are carried out using the raw sensed data received from the sensor node. Classifications based on ANN, Naive Bayes, and KNN have been adopted and tested respectively. The test results and their evaluation are presented in section V.

B. Postures Recognition with Filtered Sensed Data

As data transmission in WSNs consumes most of the energy, the lifetime of battery-powered WSNs will be prolonged if the same accuracy level can be achieved with only filtered sensed data. By studying the data from three-axis accelerometer in time domain, it has been found that these four postures feature significantly differ in amplitudes and frequencies. For example, the dynamic postures (walking and running) and the static postures (standing and kneeling) differ in their frequencies; both the two dynamic postures and the two static postures differ in their amplitudes. Making use of these features we can define the filter conditions on which the sensor node should send or not the data to the central base station. In this study, the conditions are represented by thresholds on the amplitude of the three axial accelerometer. The thresholds are chosen so that the filtered data that is transmitted to the central base station still captures the amplitude features for all four postures. In summary, the procedure for postures recognition with filtered sensed data consists of three stages as follows:

1. During the machine learning stage, raw data from the three-axis accelerometer is collected and the supervised learning is based on the raw data. In practice this is workable as the time for data collection in this stage is short and controllable and thus there is no concern about the lifetime of the BSNs.
2. The second stage is called pre-test stage. This stage is similar to the machine learning stage, i.e., subjects take the four postures respectively in a controlled period of time, and the raw data is transmitted. However, this stage is different in that the sensed data is from new subjects for test purpose. Based on the collected raw data, the thresholds for X-axis, Y-axis and Z-axis are determined based on the analysis and trials on the raw data. In practice, this is also workable with the same reason as in the machine learning stage.
3. The pre-test stage is followed by the test stage. The three-axis accelerometer senses the accelerations, and the Shimmer node filters the data based on the thresholds and transmits the filtered data to the central base station. On the central base station, the classifier recognizes the posture based on the received filtered data.

A case study with detailed recognition process and results is presented in section V.

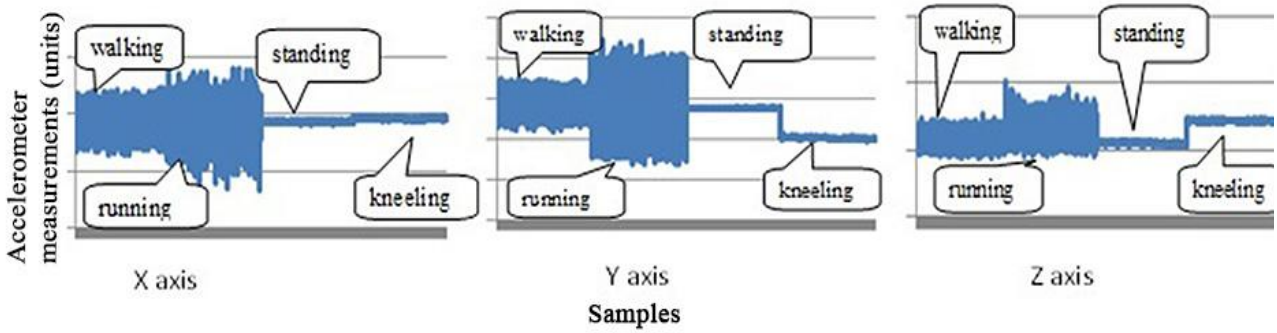


Figure 2. Sensed raw data from SHIMMER board for X, Y and Z axis for four different positions

V. RESULTS

All three classification algorithms were trained and tested with two independent sets of data collected from 12 subjects – data from six subjects with various genders and heights were collected for training and another six subjects were used for testing. In data collection we have used a Shimmer-2 sensor node with a three-axis accelerometer that was attached to the abdominal nearby the navel of the subjects. Each subject was asked to carry out four activities in a row: walking, running, standing and kneeling, with each activity being kept for 3 minutes following Dijkstra et al. [21] and Foerster et al. experiments [22]. Data packets were transmitted via Bluetooth radio to the base station where the data was stored in a Comma Separated Value (CSV) format. Raw data from the three-axis accelerometer for the four postures are shown in Figure 2, where it can be seen that these four postures feature different frequencies and amplitudes on all three axes. The software used in analyses and evaluation stage of this study was RapidMiner 5.2 [23] freely available open-source data mining and analysis system. It provides a cross platform, intuitive drag'n'drop GUI interface, which contains different machine learning and pattern recognition algorithms implemented in Java programming language.

A. Postures Recognition with Raw Sensed Data

With the raw data for both the training and testing stage, three classifiers have been constructed. The obtained results measured in terms of accuracy of these classifiers are reported in Table 1. Based on these results it can be seen that with the use of a single three-axis accelerometer, the four postures can be recognised with over 80% accuracy using all these three classifiers. Moreover we can observe that:

- ANN classifier can recognise the four postures, because in the supervised learning process, the features of the postures have been mapped to the weights of the interconnections;
- Naive Bayes classifier can recognise the four postures, because in the supervised learning process the features of the postures have been mapped to the probabilities for each axis
- KNN classifier can recognise the four postures, because in the supervised learning process, the features of the postures have been classified in terms of the distance function chosen which, in our case, is the Euclidean metric.

B. Machine Learning based Postures Recognition with Filtered Sensed Data

The procedure for postures recognition with filtered sensed data has been described in section IV.C. During the pre-test stage, raw data from a new subject is collected. After the analysis and some trials, the thresholds for X-axis, Y-axis and Z-axis, denoted as *XThreshold*, *YThreshold* and *ZThreshold* respectively, are shown in below where *avgX*, *avgY* and *avgZ* refer to the average value of the three-axis raw acceleration data from the training set respectively, and *MaxAmplitudeX*, *MaxAmplitudeY* and *MaxAmplitudeZ* refer to the maximum difference in the value of the three-axis raw acceleration data from the training set respectively:

TABLE 1. ACCURACIES OF THE CLASSIFIERS

	Raw Data	Filtered Data
ANN (500 training cycles, 0.8 learning rate, 0.0 momentum)	86.63%	86.51%
Naive Bayes	86.76%	81.67%
KNN (K=10)	82.60%	87.73%

$$XThreshold = avgX + 10\% \times \frac{MaxAmplitudeX}{2} \quad (1)$$

$$YThreshold = avgY + 50\% \times \frac{MaxAmplitudeY}{2} \quad (2)$$

$$ZThreshold = avgZ + 40\% \times \frac{MaxAmplitudeZ}{2} \quad (3)$$

With these thresholds, 1/3 of the raw data is filtered out. On the sensor node, the only change on the code is to include the condition and action part of the data filtering process, as shown below:

```
With (sensedData)
If (   (sensedData-X > XThreshold)    and
       (sensedData-Y > YThreshold)    and
       (sensedData-Z > ZThreshold))
then   (transmit (sensedData))
```

where *sensedData* refers to the data sensed by the 3-axis accelerometer.

When the filtered data is investigated, it is found that all the data for the standing posture is filtered out. So in the classifying process running on the base station, the default posture is set as 'standing' and the classifying algorithm presents as follows:

```
currentPosture = 'standing';
/* set a timer named postureTimer with the
period being 1 second */
//timeout routine for postureTimer is
{currentPosture = 'standing'}
//main algorithm
while (classifying)
{  with (receivedData)
   do
   {  currentPosture=Classifier(receivedData);
      reset (postureTimer)
   }

}
```

Table 1 illustrates the classification results for the ANN classifier, Naive Bayes classifier and KNN classifier with filtered data respectively, from which it turns out:

- With filtered data, all these classifiers are able to recognize the postures with accuracy being over 80%.
- It is noticed that when using 10-NN, the accuracy with filtered data is higher than that with raw data. This is because those filtered measurements do not represent the features of the postures, thus the relative accuracy has improved.
- 1/3 of the raw data is filtered out in this case. As data for standing posture (around 1/4 of the raw data) is filtered out, this posture is not identified by the classifiers but by the algorithm described above. For the rest of the data that is filtered out, it had little effect on the accuracy of the classifiers as it doesn't contain the information on the features of the postures.
- In practice, for different subjects, the thresholds may be different. This requires that the middleware for the sensor nodes should enable on-line parameters (such as the thresholds) re-configuration.

VI. FUTURE WORKS

As mentioned above, to make this filtered data based postures recognition practical, the underlying WSN middleware should facilitate on-line re-configuration, i.e., setting the thresholds for the three-axis accelerometer. Rule Execution and Event Distribution (REED) middleware has been developed that employs a rule-based paradigm to allow sensor networks to be programmed at run time. Setting and resetting thresholds can be implemented via updating the condition part of the rules that run on the sensor nodes. REED has been implemented on the Gumstix™ GS400K-XM, as shown in Figure 3 [24]. So far, with this platform a simple rules based test bed has been simulated on the Microchip PIC18F4520 that implements the automatic control of the temperature based on the higher and lower thresholds. In addition, a publish-subscribe service has been implemented enabling notification conditions, expressed using thresholds, to be tuned by the subscribers. Future work will focus on implementation of the ANN and Naive Bayes classifiers for real-time postures recognition with use of REED framework. The future works will include real-life testing, further study on the mechanisms of effective classification with filtered data, and application of the research results to other applications such as energy monitoring in manufacturing environment.

VII. CONCLUSION

In battery powered WSNs, data transmission consumes most of the energy. In this paper, a mechanism for postures recognition with filtered data is proposed. A WSN system consisting of one sensor device has been deployed and acceleration data is collected from real subjects. Classifiers based on ANN, Naive Bayes and KNN have been constructed using RapidMiner. Classifiers were trained with raw training data set. Postures recognitions with unfiltered and filtered testing data were carried out respectively. Results show that with appropriate thresholds, postures recognitions with filtered data achieve the same level of accuracy over 80% when compared with unfiltered data. So far the underlying WSNs middleware that enables on-line parameters such as thresholds configuration has been implemented on some sensor node platforms. The full implementation of the real-time classifier based on ANN, Naive Bayes and KNN is still in progress.

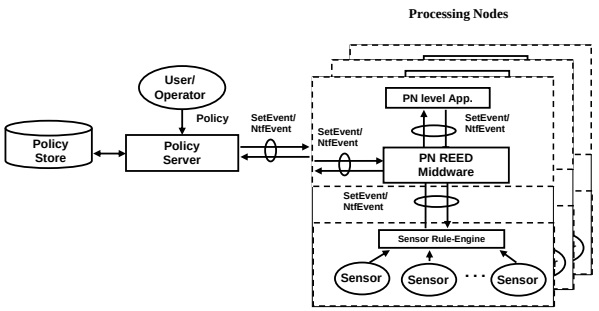


Figure 3. REED based system architecture

ACKNOWLEDGEMENT

This research was carried out as a part of the CASES project which is supported by a Marie Curie International Research Staff Exchange Scheme Fellowship within the 7th European Community Framework Programme under the grant agreement No 294931.

REFERENCES

[1] E. F. Nakamura, A. A. F. Loureiro, and A. C. Frery, "Information Fusion for Wireless Sensor Networks: Methods, Models, and Classifications,"New York,NY,USA : ACM, 2007, ACM Computing Surveys (CSUR), Vol. 39 (3).

[2] H. Ghasemzadeh, V. Loseu, and R. Jafari,"Collaborative Signal Processing for Action Recognition in Body Sensor Networks: A Distributed Classification Algorithm Using Motion Transcripts," New York, NY, USA : ACM, 2010, IPSN '10 Proceedings of thr 9th ACM/IEEE International Conference on Information Processing in Sensor Networks,pp. 244-255

[3] G. J. Pottie and W. J. Kaiser, "Wireless Integrated Network Sensors", *Communications of the ACM,* 2000, Vol. 43(5), pp. 51-58.

[4] V. Jelicic, "Power Management in Wireless Sensor Networks with High-Consuming Sensors*",* [Qualifying Doctoral Examination] Zagreb : University of Zagreb, [Online] 2011, http://www.fer.unizg.hr/_download/repository/VJelicic,KDI.pdf [retrieved: June, 2013].

[5] W. K. G. Seah, Z. A. Eu, and H-P. Tan, "Wireless Sensor Networks Powered by Ambient Energy Harvesting (WSN-HEAP)-Survey and Challenges",Aalborg : IEEE, 2009, 1st International Conference on Wireless Communication, Vehicular Technology, Information Theory and Aerospace & Electronic Systems Technology, Wireless VITAE 2009, pp. 1-5.

[6] P. Li, Y. Wen, C. Jia, and X. Li, "A Magnetoelectric Energy Harvester and Management Circuit",Washington DC, USA, 2009, PowerMEMS 2009, pp. 21-24.

[7] V. R. Challa, M. G. Prasad, and F. T. Fisher, "Towards An Autonomous Self-Tuning Vibration Energy Harvesting Device for Wireless Sensor Network Applications", IOP Publishing, 2011, Smart Materials and Structures, Vol. 20(2).

[8] D. Trabelsi, S. Mohammed, F. Chamroukhi, L. Oukhellou, and Y. Amirat, "Supervised and unsupervised classification approaches for human activity recognition using body-mounted sensors", ESANN 2012 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 25-27 April 2012,

[9] X. Long, B. Yin, and R. M. Aarts, Single accelerometer-based daily physical activity classification. In 31st Annual International Conference of the IEEE EMBS, Minneapolis, USA, September 2-6, 2009, pp. 6107-6110.

[10] M. Djuric-Jovicic, N. S. Jovicic, I. Milovanovic, S. Radovanovic, N. Kresojevic, and M. B. Popovic, "Classification of Walking Patterns in Parkinson's Disease Patients Based on Inertial Sensor Data", Belgrade : IEEE, 2010, 10th Symposium on Neural Network Applications in Electrical Engineering (NEUREL), pp. 3-6.

[11] R. Archibald and G. Fann, "Feature Selection and Classification of Hyperspectral Images with Support Vector Machines", IEEE, 2007, IEEE Geoscience and Remote Sensing Letters, Vol. 4(4), pp. 674-677.

[12] A. Godfrey, A. K. Bourke, G. M. Olaighin, P. van de Ven, and J. Nelson, "Activity classification using a single chest mounted tri-axial accelerometer", 2011, Medical Engineering and Physics, 33 (9) , pp. 1127-1135.

[13] R. R. Brooks, P. Ramanathan, and A. M. Sayeed, "Distributed Target Classification and Tracking in Sensor Networks", IEEE, 2003, Proceedings of the IEEE, Vol. 91(8), pp.1163-1171.

[14] A. Boulis, S. Ganeriwal, and M. B. Srivastava, "Aggregation in Sensor Networks: A Energy-Accuracy Trade-Off", Elsevier, 2003, Ad Hoc Networks, Vol. 1(2-3), pp. 317-331.

[15] Shimmer Research, "Shimmer", [Online] 2013, http://www.shimmer-research.com/ [retrieved: June, 2013].

[16] E. Stefanov, "Neural Networks", [Online] 2013, http://www.emilstefanov.net/Projects/NeuralNetworks.aspx [retrieved: June, 2013]

[17] E. W. Weisstein, "Conditional Probability", *Mathworld-Wolfram Web Resource*, [Online] 2013, http://mathworld.wolfram.com/ConditionalProbability.html [retrieved: June, 2013]

[18] StatSoft Inc., "Naive Bayes Classifier", *Electronic Statistics Textbook,* [Online] 2013, http://www.statsoft.com/textbook/naive-bayes-classifier/ [retrieved: June, 2013].

[19] StatSoft Inc.,"K Nearest Neighbours", *Electronic Statistics Textbook,* [Online] 2013, http://www.statsoft.com/textbook/k-nearest-neighbors/ [retrieved: June, 2013].

[20] IBM, "KNN Algorithms", *IBM SPSS Statistics Information Center*, [Online] 2011, http://publib.boulder.ibm.com/infocenter/spssstat/v20r0m0/index.jsp?topic=%2Fcom.ibm.spss.statistics.help%2Falg_knn.htm [retrieved: June, 2013].

[21] B. Dijkstra, Y. P. Kamsma, and W. Zijlstra, "Detection of Gait and Postures using a Miniaturized Triaxial Accelerometer-Based System:Accuracy in Patients With Mild to Moderate Parkinson's Disease " , Elsevier, August 2010, Archives of Physical Medicine and Rehabilitation, Vol. 91(8), pp. 1272-1277.

[22] F. Foerster, M. Smeja, and J. Fahrenberg, "Detection of Posture and Motion by Accelerometry: A Validation Study in Ambulatory Monitoring", September 1999, Computers in Human Behavior, Vol. 15(5), pp. 571-583.

[23] I. Mierswa, M. Wurst, R. Klinkenberg, M. Scholz, and T. Euler,"YALE: Rapid Prototyping for Complex Data Mining Tasks", Philadelphia, PA, USA, 2006, Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-06), pp. 935-940.

[24] X. Fei and E. H. Magill, "REED:Flexible Rule Based Programming of Wireless Sensor Networks at Runtime", Elsevier, 2012, Computer Networks, Vol. 56(14), pp. 3287-3299.

 ISBN: 978-1-61208-296-7

Sum Minimum Cost Link Algorithm for Wireless Sensor Networks

Noureddine Assad*, Brahim Elbhiri†, Sanaa El Fkihi‡, My Ahmed Faqihi‡,
Mohamed Ouadou* and Driss Aboutajdine*
*LRIT, Research Unit Associated to the CNRST (URAC 29), FSR, Med V-Agdal University, Rabat, Morocco
assad.noureddine@gmail.com
†EMSI Rabat, Morocco
‡ENSIAS Mohammed V University Souissi Rabat, Morocco

***Abstract*— A Wireless Sensor Network is a group of small sensor nodes which are used to cooperatively monitor physical or environmental conditions. Each node collects events or data from the field of interest, which seems inaccessible at the most time. By using the wireless communication capacities, sensor node sends its information to the remote base station where the end-user can access data. To deal with this, an appropriate energy-efficient routing algorithm is needed to the inherent characteristics of this kind of networks. The objective of our paper is to extend the whole network lifetime by applying two routing algorithms, one on the entire network and the other, on a special area called a hotspot, with the aim to tackle the problem of the isolation of the sink node caused by the depletion of the energy of sensor nodes surrounding it. The first protocol is based on residual energy and link distances, and aims seeking the optimal path from the sensor node transmitter to the sink node. The second protocol is based on the equitably distribution of the energy over all nodes located in the hotspot area and avoids the isolation the sink node from the rest of the network.**

Keywords-WSNs; Event detection probability; Route selection; Energy Consumption.

I. INTRODUCTION

Recent advances in technological developments at the micro-application, the radio communication and the integration of microprocessors, have made possible to develop a new range of small low-cost electronic devices. The latter may be deployed in large numbers to form an intelligent and autonomous Wireless Sensor Network (WSN) that can monitor specific areas. Each sensor node collects events or data from the field of interest, which seems the most time inaccessible, by using the wireless communication capacities. Then, sensors usually communicate with each other in a multi-hop manner and specific nodes send information to a remote Base Station (BS) or sink, where the end-user can access them [1]. Because of limitations on the energy supply, the available storage space and the computational capacity of the sensor nodes, the energy conservation is a critical challenge in a WSN [2].

The energy restrictions of sensor networks influence the design of every piece of the software running in the network. To increase energy saving, protocols might take advantage of network specific characteristics to optimize their operations, such as radius coverage, sensing coverage, event-driven operation, and node positioning [3]. Thus, to extend the whole network lifetime, protocols must interact only with nearby nodes to increase their scalability and efficiency; sensor nodes are more prone to failures and harsh communication conditions. If all sensors in a field of interest follow the nearest sink strategy, sensors around nearest sink, called hotspot, will exhaust energy early. It means that this sink is isolated from the network early and numbers of routing paths are broken. This task must be taken into account because when the sink node is isolated from the rest of the network, the WSN becomes useless [4].

The objective of this paper is to extend the whole network lifetime by applying two routing algorithms, one on the whole network and the other on a special area called a hotspot, in order to tackle the hotspot problem, the isolation of the sink node caused by the depletion of the energy of sensor nodes surrounding it.

The first protocol is based on residual energies and link distances, and aims to find the optimal path from the transmitter sensor node to the sink node.

The second protocol is based on the equitably distribution of the energy over all nodes located in the hotspot area and avoids the isolation of the sink node from the rest of the network; only nodes with residual energies greater than the network energy average will participate to relay the data parquets.

The remaining of the paper is organized as follows: In Section II, some related works are presented. Then, radio model and problem statement are reviewed in Section III. In Section IV, we describe the new proposed route algorithm, while in Section V, we present the simulation results and their analysis. Finally, conclusions are drown in Section VI.

II. RELATED WORK

In this article, we focus on the flat-based routing algorithms, in which all nodes are typically assigned equal roles or functionalities. In flat-based networks, each node typically plays the same role and sensor nodes collaborate to perform the sensing task. This is also called Multi-hop flat routing. In the next paragraphs we will give some examples of these protocols.

One of the flat-based routing techniques is SPIN (Sensor Protocols for Information via Negotiation) [5]. It is a source initiated protocol that uses a flat network structure and reactive routing. The SPIN protocol family is designed based on four basic messages that are:

 ISBN: 978-1-61208-296-7

- *ADV*: new data advertisement. This message is used by a node to inform other nodes that it has data to send; the actual data is sent only when acknowledged and requested by a node. Also, this massage contains only meta-data.
- *REQ*: request for ADV data. It is sent by the recipient to the sender node, if the recipient is interested in the actual data. Here also, it contains only meta-data.
- *DATA* : this is the actual data message.

In addition, the Directed Diffusion (DD) [6] is a data-centric (DC) and application-aware protocol in which data generated by sensor nodes is named by attribute-value pairs. In the DC protocol, data coming from different sources are combined and thus eliminating redundancy and minimizing the number of transmissions. It consists of four elements: (1) Interest: a task description which is named by a list of attribute-value pairs that describe a task, (2) Gradient: path direction, data transmission rate, (3) Data message, and (4) Reinforcement: to select a single path from multiple paths.

The MCFA (Minimum Cost Forwarding Algorithm) [7] consists in finding a path with the minimum of intermediate sensors nodes to reach the remote Sink; each node maintains the least cost estimated from itself to the Sink. Source node will forward the message to its neighbors, which is nearest to the Sink. This process repeats until the Sink is reached. Besides, the MTE (Minimum Transmission Energy) [8] proposed protocol is based on selecting route that uses the least amount of energy to transport data from the source node to the Sink. Assuming that the energy consumption is proportional to square distance between nodes, the intermediate nodes, which operate as routers, are chosen for minimizing the sum of squared distances over the path. Thus, if a node A transmits data to a node C, the node B will participate to the route only if (d_{xy} is the distance between x and y):

$$d_{AC}^2 > d_{AB}^2 + d_{BC}^2 \tag{1}$$

Furthermore, xMREPsum (Sum Maximum Residual Energy Path) [9] is another route technique. Here, the authors propose a cost function that represents the reciprocal of the residual energy at node N_i after the route will have been used by a packet, and uses an exponential parameter P. This function is given by the next equation:

$$f(c_{ij}) = \left(\frac{1}{E_{ri} - e_{ij}}\right)^P \tag{2}$$

where E_{ri} is the residual energy of the node N_i and e_{ij} is the energy expenditure per bit transmission across the link (N_i, N_j). This last energy is defined as follows:

$$e_{ij} = max\left(0.01, \frac{d_{ij}}{E_i}\right) \tag{3}$$

where E_i is the initial energy of the node i, and d_{ij} is the distance between the two nodes N_i and N_j.
Then, the shortest path is obtained by using the conventional BellmanFord [10] equation given by:

$$D_i^{n+1} = min\left(min\left(f\left(c_{ij}\right) + D_j^n\right), D_i^n\right) \tag{4}$$

III. RADIO MODEL AND PROBLEM STATEMENT

We consider a simple model for the radio hardware energy dissipation where the transmitter dissipates energy to run the radio electronics and the power amplifier, and the receiver dissipates energy to run the radio electronics. This model is the one used in [11] and illustrated in Fig. 1.

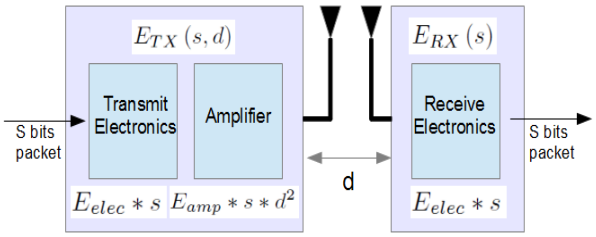


Fig. 1. Radio energy dissipation model.

The transmitted and received energy costs for the transfer of an s bit data message between two nodes separated by a distance d, is given by the following equations:

$$E(s,d) = E_{TX}(s,d) + E_{RX}(s) \tag{5}$$
$$E_{TX}(s,d) = E_{elec} * s + E_{efs} * s * d^2 if d \leq d_0 \tag{6}$$
$$E_{TX}(s,d) = E_{elec} * s + E_{amp} * s * d^4 if d > d_0 \tag{7}$$
$$E_{RX}(s) = E_{elec} * s \tag{8}$$

where E_{elec} is the energy dissipated per bit to run the transmitter E_{TX} or the receiver E_{RX} circuit. The E_{elec} depends on many factors such as the digital coding, the modulation, and the spreading of the signal. E_{amp} is the transmitter amplifier while E_{efs} is the transmitter in free space, d is the distance between the sender and the receiver, and $d_0 = \sqrt{\frac{E_{efs}}{E_{amp}}}$ is the minimal distance. As a conclusion, the total energy consumption E_k in the selected route P_k which source node sends s bits packet to the Sink is as below:

$$E_k = \sum_{i=0}^{D-1} E_{i,i+1} \tag{9}$$

where D is the number of nodes along the route of source destination pair.

The basic function of a routing algorithm is to select the path from a set of paths, which is the most effective on the basis of specific criteria [12][13]. Intuitively, to maximize the lifetime of WSNs, the path that achieves the minimum energy consumption while ensuring energy field between individual nodes must be used [14][15]. Flat multi-hop routing algorithms are excellent in terms of their capability of using power-aware metrics to choose minimum power consuming paths. Our proposed routing protocol is combined between two criteria, the first is the shortest path, because data from sensor networks are usually time sensitive; so it is important to receive data as soon as possible; the second criteria is the residual energy of nodes, which distributes equitably the energy over all nodes.

 ISBN: 978-1-61208-296-7

The isolation of the sink node is caused by the depletion of the energy of sensor nodes surrounding it (the hotspot problem). It means that this sink is isolated from network early and numbers of routing paths are broken, so the WSNs becomes useless [16]. To avoid the isolation of the sink node, we aim to suggest a new strategy which distributes equitably the energy over all nodes of a WSN.

Another challenge of the WSNs is the efficient deployment of the required coverage [17]. Specifically, given a monitoring region, how can we guarantee that every point in the region is covered by the required number of sensors? In other words, we need to recognize which areas are covered by enough sensors. This problem is challenging due to the limitations of wireless sensors. In the next section, we will define sensing coverage and will formulate the maximum sensing coverage region problem.

IV. PROPOSED ROUTE ALGORITHM

To reduce the energy consumption in WSNs, we propose a protocol architecture to ensure good routing of data between nodes and the BS. We adopt a sensing model probability in dense networks. A sensor can detect any events located in the area sensing, so we can identify the redundant nodes that will change their operating modes between sleep and active ones. In this section we will give more details of the different components of our proposal MCLsum (Sum Minimum Cost Link) algorithm.

A. Event detection probability

WSNs are characterized by random network topology where all nodes are randomly deployed in field of interest. Our objective in this section is to define the optimal number of the nodes used in an area and sensing coverage. To achieve this objective, we study the sensing model probability that is generally coupled closely with the specific sensor application and the type of sensor device used [18][19]. We adopt a sensing model probability where a sensor can detect any events located in the area sensing. All sensor nodes are assumed to be homogeneous, and have the same sensing coverage (R_s).

For a uniformly distributed sensor network with node density λ [20], we denote the number of sensors covering the location by n. The sensing model probability follows a Poisson distribution as given by the next formula:

$$P(n=k) = \frac{(S\lambda)^k}{k!} e^{(-S\lambda)} \tag{10}$$

P is the probability that an event can be immediately detected by k sensors located within the area of interest, each sensor covering an area of $S = \pi R_s^2$. Hence, the previous probability can further be represented as:

$$P(n=k) = \frac{(\lambda\pi R_s^2)^k}{k!} e^{(-\lambda\pi R_s^2)} \tag{11}$$

Therefore, the probability that no sensor in an area of interest can detect an event is $\overline{P} = e^{(-\lambda\pi R_s^2)}$. Then, the complement of $\overline{P}$ is the probability that there is at least one sensor which detects an event. This sensing model probability can be represented as:

$$P = 1 - \overline{P} = 1 - e^{(-\lambda\pi R_s^2)} \tag{12}$$

P is determined by the node density and the sensing range, while increasing the sensing coverage and the number of nodes, the detection probability increases and all events that happen in the network, will be covered. The optimal values that must be considered to totally cover an area are obtained by the simulation of the formula event detection probability.

B. Availability of nodes and redundant nodes

We can efficiently reduce the energy consumption in most of the WSNs applications, and extend the whole network lifetime, when sensor power can be put on/off periodically. Thus, it is appropriate to take into account the node availability rate p in our analysis. Each sensor node can decide whether to become active with probability p or move to the sleep mode with probability $1-p$ witch means to be off in every sensing period. Thus, the probability that the event can be immediately detected once it appears in an homogeneous WSNs with node density λ, sensing range R_s and node availability p, can be given determined by:

$$P = 1 - \overline{P} = 1 - e^{(-p\lambda\pi R_s^2)} \tag{13}$$

In a dense network, the sensing areas of different nodes may be similar to their neighbor nodes and nodes will transmit redundant information. This will increase the total energy consumption of the WSN. Therefore, it is important to place or select the effective number of sensor nodes to cover the same monitored area as much as possible without diminishing the overall field coverage. Thus, we must identifies the redundant nodes denoted by probability $1-p$ in a dense networks and changes their operating mode between sleep and active modes.

1) The overlapped area and redundant nodes in WSNs: Consider two nodes N_i and N_j location at (x_{Ni}, y_{Ni}) and (x_{Nj}, y_{Nj}) respectively. The node N_i covers the S_{Ni} area and the other node covers the S_{Nj} area. Let us note d_{ij} the distance between nodes N_i and N_j.

The overlapped area S_{ol} of nodes N_i and N_j is defined as the intersection region between S_{Ni} and S_{Nj} areas, that is formulated by the next equation:

$$S_{ol} = S_{Ni} \cap S_{Nj} \tag{14}$$

An example of S_{ol} is represented in Fig. 2.

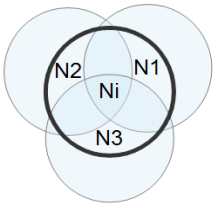


Fig. 2. Example of redundant node N_i totally overlapped by nodes N_1, N_2, and N_3.

As in [17], S_{ol} can obtained by the follow result:

$$S_{ol} = R_s^2 \times \left(\theta - \frac{d_{ij}}{R_s} \sqrt{1 - \left(\frac{d_{ij}}{2R_s} \right)^2} \right) \quad (15)$$

with θ is the central angle of the overlapped area given by the next equation:

$$\theta = \arccos \left(\frac{d_{ij}}{2R_s} \right) \quad (16)$$

We can note that each point located in S_{ol} can be covered by both N_i and N_j nodes.

2) Operation modes of nodes: Due to the fact that the deployment of node in the network is random, many redundant nodes may be detected [21]. As discussed above, a redundant node is a node where its entire sensing region is covered by the sensing region of other nodes. Hence, we run our algorithm to determinate this kind of nodes. Assume that each node knows positions of its neighbours, it is possible that all of them have the capability to define their state (redundant or not). If the node is redundant, it powering-off both the sensing and the communication units. Here, we note that the node is operating in the sleep mode, but from time to time it powering-on only the radio module to listen if data is coming from its neighbours during a listening time T_{listen}. Consequently, the remaining energy is saved and the whole network lifetime is extended. But, how long should a redundant node remain in a sleep mode? To response to this question, we apply the following approach based on the minimum residual energy threshold parameter. Node N_i participating to cover a part of one redundant node and having residual energy E_{ri} greater than threshold Θ_{Eng}, will broadcast continually a short message during a period $T_{broadcast}$ with a radio range equal to R_s. The objective of these transmissions is to inform the redundant node that its remaining energy will be exhausted soon. Indeed, the sufficient and necessary conditions to wake-up the redundant node are:

$$T_{broadcast} > T_{listen} \quad (17)$$

$$E_{ri} > \Theta_{Eng} \quad (18)$$

Thus, the redundant nodes verifying the two last conditions, can switch from the sleep mode to the active one. Therefore, a large part of energy will be saved and the redundant transmissions will be efficiently reduced. Consequently, the whole network lifetime will be extended.

C. Routing inside Hotspot area

We suggest a new protocol to select nodes that will participate to the data transmission path in the hotspot area. It is based on a strategy which distributes equitably the energy over all nodes located in the hotspot area and avoids the sink node isolation from the rest of the network. Sink defines the radius of hotspot area by broadcasting a short message, and requesting the residual energy for each node located in hotspot area. In order to compute the energy average of the hotspot area, each node sends a short message to the sink containing the residual energy; only nodes with residual energy greater than the network energy average will be selected by the Sink to participate to relay the data packets, as expressed by the next equation:

$$Er_i \geq Emean : i \in B \quad (19)$$

where Er_i is the residual energy of the node i, $Emean$ is the energy average of hotspot area, and B is the node set located in hotspot area.
The diagram in Fig. 3 shows more details and resumes all these steps.

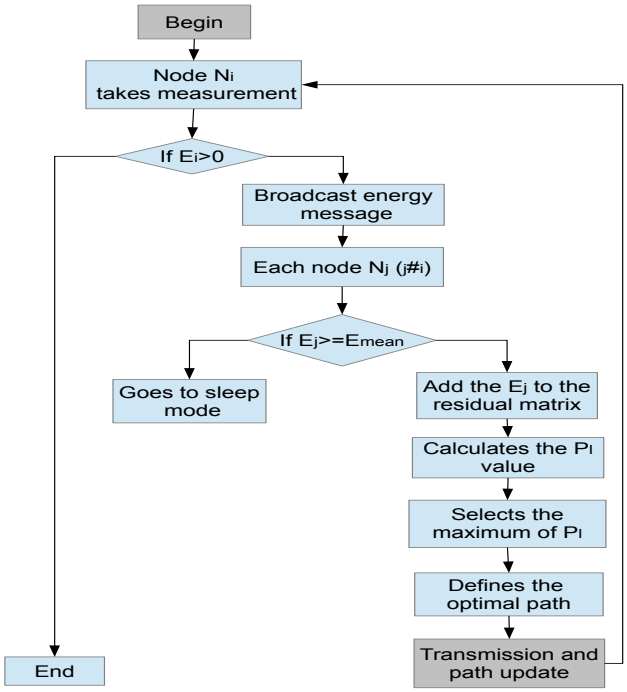


Fig. 3. Routing inside Hotspot area flow chart.

D. Selecting an optimal route

It consists to select paths with maximum residual energy and squared distances. So, energy consumption is balanced among all nodes with the aim to minimize the sum of squared distances over the path. More specifically, cost function that represents the link c_{ij} is:

$$f(c_{ij}) = \frac{d^2}{Er_j} = \frac{\mid N_i - N_j \mid^2}{Er_j} \quad (20)$$

where Er_j is the residual energy of the node j and the distance $\mid N_i - N_j \mid$ is defined as the length of the line segment connecting N_i and N_j; this Euclidean distance must be less than the coverage radius R_c of node i. The total cost for a route l of length D is computed by the following equation:

$$P_l = \sum_{i=0}^{D-1} f(c_{i,i+1}) = \sum_{i=0}^{D-1} \frac{\mid N_{i+1} - N_i \mid^2}{Er_{i+1}} \quad (21)$$

 ISBN: 978-1-61208-296-7

Using the Dijksta algorithm, the optimal route with the minimum P_l can be fixed. As a conclusion, the selected route P_k is the one that satisfies the following property:

$$P_k = min\{P_l : l \in A\} \quad (22)$$

where A is the set of all the possible routes. Our proposed scheme, routing inside hotspot area is shown in Fig. 4.

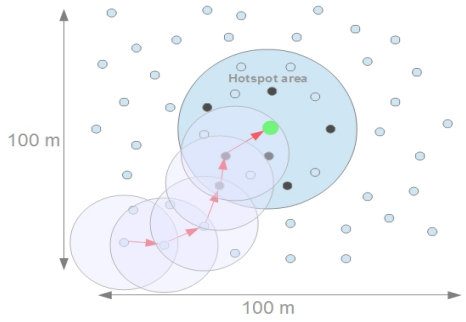


Fig. 4. Wireless sensor networks.

V. SIMULATION RESULTS

In this section, we evaluate the performance of our architecture protocol using MATLAB software. We consider a random wireless sensor network composed of n static sensor nodes and one sink node. These static sensor nodes are independent and distributed uniformly on an area $100m \times 100m$. Each sensor node sends its acquired data to the sink node located at the center of the sensor field. Moreover, we run our simulation using the radio model presented in Fig. 1 with simulation parameters shown in Table I.

TABLE I
SIMULATION PARAMETERS

NODES	100
INITIAL ENERGY E_r	5 J/node
TRANSMITTER ELECTRONICS E_{elec}	50 nJ/bit
E_{efs}	10 nJ/bit
E_{amp}	100 nJ/bit
DEAD NODES	$<$ 0.1 J
DATA PACKET SIZE	4000 bytes
ENERGY MESSAGE SIZE	200 bytes

A. Optimal node number with sensing coverage for an area

The results illustrated in Fig. 5 shows that the event detection probability P is determined by the node density and the sensing range. Event detection may need a large sensing range or a high node density, thus increasing the WSN deployment cost. Fig. 5(a) shows the simulation results of the formula event detection probability. On the other hand, we can obtain the same results given in Fig. 5(b). By running simulations, we change the node density by varying two parameters, the sensing range R_s from $0m$ to $20m$, and the node number n from 1 to 100. We assume that an event occurs randomly in a given location on the network and we compute the probability to detect this event. For given values of R_s and n, we run 10000 simulations and we calculate the percentage of detection. We can note that if we increase the lasts parameters the probability to cover an event happens in the network increases to. However, for a given value of R_s or n we can see that this percentage of coverage is constant and attend 100%. Indeed, increasing more n or R_s will not affect the robustness of detection. Consequently, for a given value of R_s, we can find the optimal node number which can be deployed to cover efficiently the controlled region and reciprocally. This node number and sensing range will be the optimal values, which must be used to totally cover an area.

B. Node availability

We set the node number to 100 and the sensing coverage to 20. we run our algorithm to determinate redundant nodes.

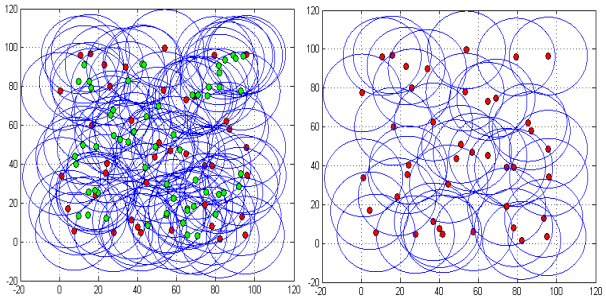


Fig. 6. Redundant nodes for a $100m \times 100m$ network area.

Fig. 6 shows the redundant node location determined in the WSN area. We note that there are approximately 35 redundant nodes expressed by the green color dispersed over the network. Moreover, we see that only 65 nodes can cover efficiently the total region. So, 65 is the optimal node number which can be deployed to cover the region.

We can express the efficient number of nodes which can be deployed to cover the area of interest by node availability rate p as a variable in event detection probability P.

$$P = 1 - e^{(-p\lambda\pi R_s^2)} \quad (23)$$

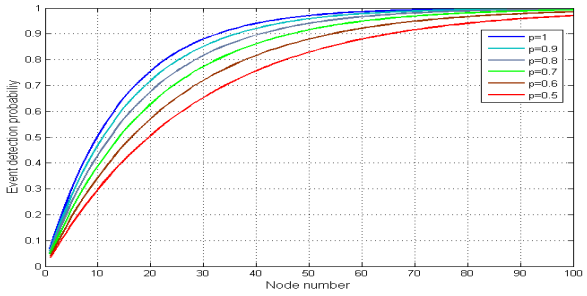
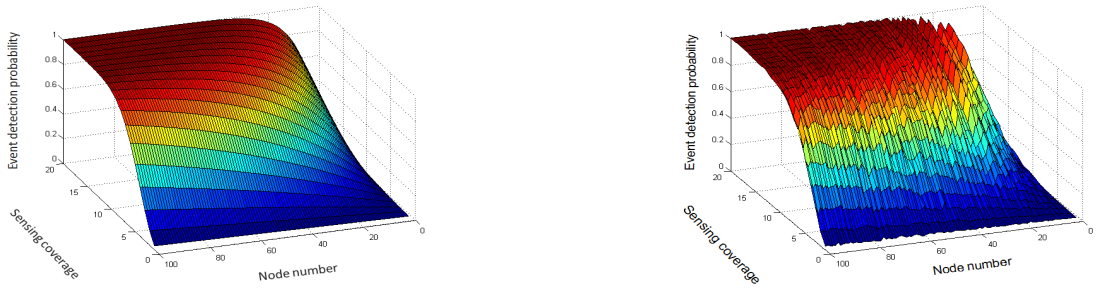


Fig. 7. Node availability.



(a) (b)

Fig. 5. Event detection probability, (a) Sensing model probability (b) Simulation.

Fig. 7 shows the curves of event detection probability model according to different values of node availability rate. It is obvious that if node availability rate p increases the event detection probability P increases too.

In the next subsection, we will evaluate the performance of our routing protocol MCLsum by taking into account that redundant nodes switch from sleep mode to active one, and vice versa.

C. Performance evaluation

We give the comparisons between MCLsum, the MTE and the xMREPsum algorithms. The evaluations are discussed in terms of energy stored, number of nodes still alive, and number of messages received at the sink.

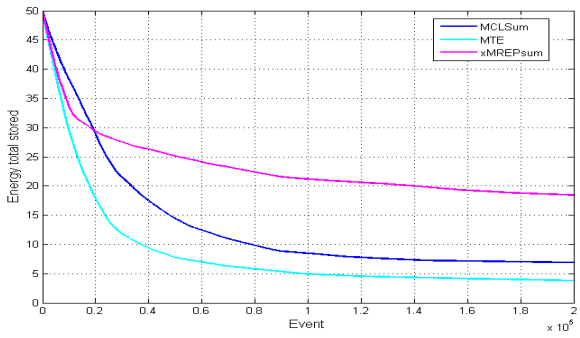


Fig. 8. Total energy stored (Joule).

In each round, we assume that an event occurs randomly in a given location on the network. During the simulation, the network nodes die randomly in the supervised area. We then give the simulation results of the three compared approaches.

Fig. 8 gives the total network remaining energy in every sensing round. The network remaining energy decreases rapidly in the Direct transmission and xMREPsum protocols.

We can see that, in the $0.1 * 10^6$ first events, approximately 45% of the total network energy is consumed in the MTE and 35% in xMREPsum protocol. Whereas, the MCLsum consumed only 20% of the total energy of the network. From the $0.2 * 10^6$ first events, rate event detection for protocol xMREPsum is reduced compared to two other protocols MTE and MCLsum, which explains low energy consumption.

As shown in Fig. 9, we represent the number of data messages that are correctly received by BS. We can notice that the total number of these messages is substantial for the MCLsum protocol. This means that MCLsum is more efficient than MTE and xMREPsum protocols.

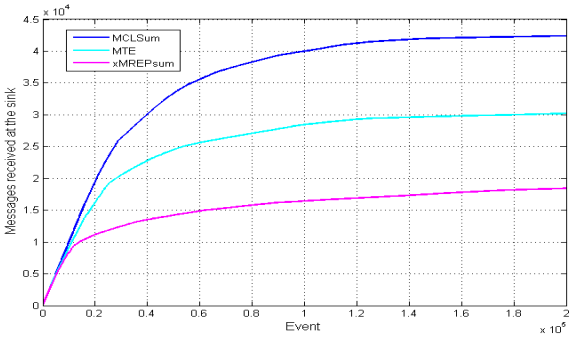


Fig. 9. Number of data messages correctly received at the Sink.

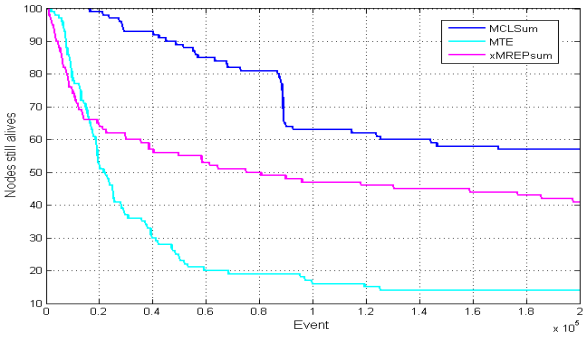


Fig. 10. Number of nodes still alive.

Fig. 10 shows the variation in the total number of sensors that are still alive. We observe that in MTE protocol, the first node die appears in the 1800th event, in xMERPsum it happens in the 900th event and in MCLsum it comes in 16300th event. So, our protocol increases the whole network lifetime. Moreover, our MCLsum protocol is approximately 2x better than MTE in term of node death, and better about 3x than xMREPsum. Taking in consideration the residual

energy, square distance between nodes, and average network energy in hotspot area, our protocol presents more efficiency and more robustness. Consequently, the MCLsum enlarges substantially the lifetime of the whole network related to the others algorithms.

VI. Conclusion

In this paper, we presented a novel routing protocol for WSNs by taking into account the hotspot problem. The proposal provides better solution for routing protocol of WSNs. It is based on residual energy and link distances of nodes, and aims to seek the optimal path from a transmitter sensor node to a sink node. The hotspot problem is solved thanks to our strategy which equitably distributes the energy over all nodes located in the hotspot area and avoids the isolation of the sink node from the rest of the network. The simulation results show that, compared to the MTE and the xMREPsum protocols, best results are obtained by the MCLSum protocol. Thus the whole WSNs lifetime is significantly extended by the last approach.

References

[1] I. F. Akyildiz, T. Melodia, and K. R. Chowdhury, "A survey on wireless multimedia sensor networks," *Computer networks*, vol. 51, no. 4, pp. 921–960, 2007.

[2] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "A survey on sensor networks," *Communications magazine, IEEE*, vol. 40, no. 8, pp. 102–114, 2002.

[3] Y. Wang, "Topology control for wireless sensor networks," in *Wireless Sensor Networks and Applications*. Springer, 2008, pp. 113–147.

[4] B. Nazir and H. Hasbullah, "Mobile sink based routing protocol (msrp) for prolonging network lifetime in clustered wireless sensor network," in *Computer Applications and Industrial Electronics (ICCAIE), 2010 International Conference on*. IEEE, 2010, pp. 624–629.

[5] E. Woodrow and W. Heinzelman, "Spin-it: a data centric routing protocol for image retrieval in wireless networks," in *Image Processing. 2002. Proceedings. 2002 International Conference on*, vol. 3. IEEE, 2002, pp. 913–916.

[6] C. Intanagonwiwat, R. Govindan, D. Estrin, J. Heidemann, and F. Silva, "Directed diffusion for wireless sensor networking," *Networking, IEEE/ACM Transactions on*, vol. 11, no. 1, pp. 2–16, 2003.

[7] J. N. Al-Karaki and A. E. Kamal, "Routing techniques in wireless sensor networks: a survey," *Wireless Communications, IEEE*, vol. 11, no. 6, pp. 6–28, 2004.

[8] M. Ettus, "System capacity, latency, and power consumption in multihop-routed ss-cdma wireless networks," in *Radio and Wireless Conference, 1998. RAWCON 98. 1998 IEEE*. IEEE, 1998, pp. 55–58.

[9] D. J. Vergados, N. A. Pantazis, and D. D. Vergados, "Energy-efficient route selection strategies for wireless sensor networks," *Mobile Networks and Applications*, vol. 13, no. 3-4, pp. 285–296, 2008.

[10] G. L. Nemhauser and L. A. Wolsey, *Integer and combinatorial optimization*. Wiley New York, 1988, vol. 18.

[11] T. S. Rappaport *et al.*, *Wireless communications: principles and practice*. Prentice Hall PTR New Jersey, 1996, vol. 2.

[12] Y. Liu and Z. Wang, "Maximizing energy utilization routing scheme in wireless sensor networks based on minimum hops algorithm," *Computers & Electrical Engineering*, 2012.

[13] M. Patil and R. C. Biradar, "A survey on routing protocols in wireless sensor networks," in *Networks (ICON), 2012 18th IEEE International Conference on*. IEEE, 2012, pp. 86–91.

[14] M. Noori and M. Ardakani, "Characterizing the traffic distribution in linear wireless sensor networks," *Communications Letters, IEEE*, vol. 12, no. 8, pp. 554–556, 2008.

[15] N. Bartolini, T. Calamoneri, A. Massini, and S. Silvestri, "On adaptive density deployment to mitigate the sink-hole problem in mobile sensor networks," *Mobile Networks and Applications*, vol. 16, no. 1, pp. 134–145, 2011.

[16] S.-T. Cheng and T.-Y. Chang, "An adaptive learning scheme for load balancing with zone partition in multi-sink wireless sensor network," *Expert Systems with Applications*, vol. 39, no. 10, pp. 9427–9434, 2012.

[17] V. Tran-Quang and T. Miyoshi, "A novel gossip-based sensing coverage algorithm for dense wireless sensor networks," *Computer Networks*, vol. 53, no. 13, pp. 2275–2287, 2009.

[18] C.-Y. Chong and S. P. Kumar, "Sensor networks: evolution, opportunities, and challenges," *Proceedings of the IEEE*, vol. 91, no. 8, pp. 1247–1256, 2003.

[19] J. Jia, X. Wu, J. Chen, and X. Wang, "Exploiting sensor redistribution for eliminating the energy hole problem in mobile sensor networks," *EURASIP Journal on Wireless Communications and Networking*, vol. 2012, no. 1, pp. 1–11, 2012.

[20] Y. Wang, X. Wang, B. Xie, D. Wang, and D. P. Agrawal, "Intrusion detection in homogeneous and heterogeneous wireless sensor networks," *Mobile Computing, IEEE Transactions on*, vol. 7, no. 6, pp. 698–711, 2008.

[21] I. Demirkol, C. Ersoy, and F. Alagoz, "Mac protocols for wireless sensor networks: a survey," *Communications Magazine, IEEE*, vol. 44, no. 4, pp. 115–121, 2006.

 ISBN: 978-1-61208-296-7

IEEE802.15.4 Performance in Various WSNs Applications

Marwa Salayma, Wail Mardini, Yaser Khamayseh, Muneer Bani Yassein
Department of Computer Science
Jordan University of Science and Technology (JUST)
Irbid, Jordan
{mksalayma, mardini, yaser, masadeh}@just.edu.jo

***Abstract*— IEEE802.15.4 is a standard proposed to support physical and MAC layers for low data rate Wireless Sensor Network (WSN) applications. It provides features that allow diverse WSN applications to work at reduced energy consumption in low cost. These applications can be classified according to the arrival rate of the surrounding phenomena and according to the duty cycle. This paper investigates three WSN applications with different arrival rates and different duty cycles (100% and 50%). This paper performs intensive simulation analysis on seven network scenarios with different number of nodes. The scenarios are evaluated in terms of energy consumption, average end-to-end delay and throughput. The obtained results reveal that, irrespective of application type, both average delay and throughput behaviors vary directly, whereas energy consumption varies inversely. Results also revealed that as the duty cycle increases, both average delay and throughput improve. Improving one of the two metrics, by just increasing the duty cycle, assures enhancing the other metric, accordingly.**

Keywords-*Wireless sensor networks; IEEE802.15.4; Beacon enabled; Arrival rate; Superframe structure; Energy consumption.*

I. INTRODUCTION

Wired sensors are replaced with wireless ones to form Wireless Sensor Networks (WSNs). A WSN is a collection of sensing devices that communicate with each other and with the surrounding environment via wireless links [1][2]. The main source of power for the sensors is batteries. Due to the limited nature of batteries, it is essential to design sensor nodes to operate for months or even years [2]. Energy conservation is the main interest of literature studies in WSNs, which focus on designing WSN energy efficient algorithms and standards, one of which is the IEEE802.15.4 standard [3].

IEEE802.15.4 standard supports both physical and Media Access Control (MAC) layers. IEEE802.15.4 MAC supports two types of devices; Full Functional Devices (FFDs) and Reduced Functional Device (RFDs) which differ in their capabilities. FFDs can act as a coordinator or as a sink node and is typically referred to as PAN coordinator (PANc). On the other hand, RFD acts only as an ordinary end device [4-6]. Nodes with different types that follow the standard can communicate with each other's forming two types of topologies: star and peer to peer. Peer to peer topology can be classified as either a mesh or a cluster tree topology [6-8].

IEEE802.15.4 standard supports three different Radio Frequency (RF) bands [4]. The most popular RF band is the unlicensed 2.4 GHz. This band offers high data and it consumes less power [4]. IEEE802.15.4 MAC operates either in beacon enabled or beaconless modes. In beacon enabled mode, FFD devices broadcasts beacon frame regularly in order to identify itself so that other nodes can recognize their master and start associating with it. Beacon frame includes information that enables nodes synchronize with each other when they need to access the channel [9][10]. Furthermore, beacon frame includes information that indicates whether there is any pending data for nodes. As coordinator sends a beacon frame and follows it by another one, the time between those two frames is referred to as the Beacon Interval (BI) and it is divided virtually into 16 equal sized slots. BI duration can be specified through Beacon Order parameter (BO) according to the following formula [9]:

$$BI = aBaseSuperframeDuration * 2^{BO} \quad (1)$$

Nodes can use the channel during the whole BI period or can sleep for some time. The parameter used to specify this time is the Superframe Order (SO) according to the following formula [9]:

$$SD = aBaseSuperframeDuration * 2^{SO} \quad (2)$$
$$\text{where } 0 \le SO \le BO \le 14$$

aBaseSuperframeSuration value depends on the slot duration according to the following formula:

$$aBaseSuperframeDuration = aBaseslotDuration * total\ number\ of\ slots \quad (3)$$

Typically, time durations are expressed in terms of general time unit that is a symbol. The value of one symbol in seconds depends on the chosen RF band. In the 2.4 Ghz RF band, a symbol equals to 15.36 ms. These concepts can be summarized through one general concept namely, the duty cycle (D). It is the percentage of time the node is awake in the BI, and mathematically is expressed as follows [9][10]:

$$D = SD/BI * 100\% \quad (4)$$

 ISBN: 978-1-61208-296-7

When a node needs to access the medium, it has to locate the beginning of the next time slot to compete for the channel, thus, it follows the contention based algorithm followed by the standard, that is, the slotted Carrier Sense Multiple Access/ Collision Avoidance algorithm (CSMA/CA); this time portion is referred to as Contention Access Period (CAP) [9][10]. Furthermore, the standard empowers the PANc with the authority to assign some slots excessively for some nodes during which they can utilize the channel alone. These time slots are referred to as Guaranteed Time Slots (GTS). The optional period, which includes those slots, is referred as Contention Free period (CFP) and can include maximum of seven GTS. CAP and the optional CFP together are referred to as the Active Period, and its duration often referred to as SuperFrame Duration (SD) [9][10]. More precisely, every time the node needs to access the channel, it needs to locate the boundary of what is called the slotted/un-slotted CSMA backoff period. Thus, the standard deals with time in terms of backoff period unit which in beacon enabled mode is aligned between the slot boundaries and indicated through aUnitBackoffPeriod which equals to 20 symbols or 0.32 ms [10]. The length of each period discussed previously is assigned through beacon frame which is transmitted in the first time slot (slot 0) [4].

Improving the beacon enabled standard performance is directly related to the chosen BO and SO values. These parameters need to be chosen carefully to fully utilize the scarce resources of WSNs. How to decide the optimal BO and SO values that achieve the best performance is an application related issue. WSN applications are diverse and can be classified according to the percentage of duty cycle they work through as well as their packets arrival rate. For example, an application may have packets ready for transmission every second but need to be active for 30 minutes, and hence, shall sleep for the other 30 minutes in an hour. Some applications may work through low arrival rate such as 0.1 second, thus, can be classified as very active applications. Other applications may work through high arrival rates such as 2 seconds and classified as inactive applications, while many applications wok at low constant arrival rate which may increase suddenly according to some criteria. Hence, each application has its own special case that has much to do in the decision of BO and SO, keeping in mind that the optimal building block of any network topology consists of seven nodes (piconet).

Simulation analysis conducted in [11] focus on understanding the behavior of beacon enabled PAN according to all (BO, SO) possible combinations. However, results achieved consider one type of application with 1 second arrival rate. Moreover, results indicate that it may be impossible to achieve high PAN performance in term of all metrics. Hence, to generalize this analysis, we need to study different arrival rates for different duty cycles to improve the standard performance in terms of the three metrics irrespective of application type, which is the main goal of this paper.

This paper is organized as follows. Section II summarizes some of the literature work which is closely related to the paper topic, while Section III illustrates the followed methodology, whereas Section IV clarifies and discusses the simulation results achieved. Section V concludes the paper and offers some ideas for future work.

II. RELATED WORK

Several studies were conducted to analyse the performance of the IEEE802.15.4 standard to improve energy consumption. Many researches were proposed to tune the standard parameters to achieve the best performance of the network.

Salayma et al. [11] investigates the standard beacon enabled mode behaviour through intensive simulation applied on seven PAN scenarios. It categorizes different WSN application according to duty cycle. It finds out the optimal range of (BO, SO) combinations for 1s arrival rate applications in terms of energy consumption, average end-to-end delay and throughput. Moreover, it proposes an adaptive algorithm that outperforms the original MAC irrespective of the duty cycle. The proposed adaptive algorithm converges to the network current performance and improves network performance accordingly. However, such a work considers only Constant Bit Rate (CBR) applications with 1 second interval and results revealed may not reflect other WSN applications behavior.

Neugebauer et al. [12] proposes an algorithm that reconfigures only the BO parameter of the IEEE 802.15.4 superframe structure. The Beacon Order Adaptive Algorithm (BOAA) considers star topology. Changing BO depends on the inter-arrival rate which reflects the frequency of communication. Adjusting the value of BO changes the length of the duty cycle as a result of the dynamic changing of the beacon interval. Experimental results show that increasing the value of BO contributes to saving power [12]. However, this power saving improvement is at the expense of sharp increase in the delay, because increasing BO would cause nodes to wait for more time for the next beacon. Throughput was not taken into consideration. This makes BOAA suitable only for simple applications.

Shu et al. [6] proposes an optimization scheme, in order to achieve the minimum energy consumption under the packet delivery reliability constraints. The objective function was achieved after finding the optimal values for the two decision variables: BO and SO. The obtained results in [6] show that, for a network with number of nodes varies from 5 to 35, the optimal BO value is 7. On the other hand, for a network with number of nodes is less than 15, the optimal SO value is 1. And for the other cases, the optimal SO value is 2. However, choosing optimal values for BO and SO depends much on the chosen quality of service constraints.

In [13], the performance of the slotted CSMA/CA is investigated by studying the effects of SO, BO and Backoff Exponent (BE). However, similar study is conducted in [3] with different criteria, such as, number of nodes and data

 ISBN: 978-1-61208-296-7

frame size. Simulation experiments were conducted for 13 different values of BO and SO, all with a duty cycle of 100%. The best range of offered load that achieves the optimal trade-off between throughput and average delay utility was found to be between 35% and 60%. This study does not study the network with sleep period enabled.

In [14], the IEEE 802.15.4 standard performance is evaluated in terms of throughput and packet delivery ratio. The study focuses on the Quality of Service QoS for real time sensor applications and provides an enhancement to the current IEEE 802.15.4 beacon enabled standard by dynamically allocating the already existed GTS. The standard performance metrics were evaluated through varying both BO and SO, while preserving the dynamically allocated one GTS. However, the study considers the SO and BO values up to 6, due to association latency that may result from choosing higher values which is not sufficient for WSN applications.

In [15], the performance of beacon enabled mode IEEE 802.15.4 is evaluated in term of energy consumption in a large scale clustered tree network. Analysis of the IEEE 802.15.4 MAC were performed on real Zigbee nodes applied on home network areas by varying BO values between 6 and 10, while fixing SO value to 0. Results reveal that power consumption decreases by increasing BO to some value (approximately 10) after which it increases. However, the study considers only very low duty cycles, due to the small fraction of CAP and does not consider the effect of SO on the standard performance at all.

In [16], IEEE 802.15.4 standard performance is investigated in terms of throughput, energy consumption and reliability by applying the standard on ideal and non-ideal star topologies. The study focuses is on changing number of nodes, while varying some of IEEE 802.15.4 standard configurations, such as, the availability of synchronization, BO and SO. According to the achieved results, some recommendations were suggested that aid in configuring the standard, configuring applications that follow the standard and how to improve the standard. However, such recommendations can only be considered when applying the standard on the same tested topologies.

III. METHODOLGY AND PROBLEM ANALYSIS

Simulation work conducted by Salayma et al. [11] discusses the standard behavior in terms of total energy consumption, average end-to-end delay and throughput for seven PAN scenarios with different number of nodes. The work in [11] considers one type of applications, and the obtained results may not reflect other applications behavior. Moreover, results indicate that it may be impossible to achieve high PAN performance in term of all metrics. In other words, enhancing some metrics performance may be sacrificed in order to improve other metrics. Therefore, to generalize our analysis, our methodology considers other WSN CBR applications. It investigates active applications which receive packets very frequently, as well as, inactive applications that work through relatively low arrival rates. This paper examines different applications; as it studies the beacon enabled mode behavior which works through different arrival rates at different duty cycles.

Hence, as a first step, we need to examine what would happen as BO and SO increase or decrease by 1 in order to realize why, how much and which metric performance is needed to be enhanced adaptively. Increasing and decreasing BO and SO by 1, means that network shall operate through 50% and 100% duty cycles. WSNs applications performance is analyzed in terms of the three metrics: total energy consumption, average end-to-end delay and throughput. Applications differ in packets arrival rate which are 0.1s, 1s and 2s. A previously specified (BO, SO) combination is allowed to change dynamically at a specified time for each application in order to study the effect of increasing or decreasing of BO or SO or both by 1.

We will study the effect of choosing the same optimal combinations achieved for 1s arrival rate application on the other two applications. According to analysis results achieved in [11], it is revealed that (7, 7) is one of the combinations that relatively achieve good performance for the 1s arrival rate application, it is chosen as a reference point to analyze the other two applications behavior. Hence, after 500s, four scenarios are examined, which are increasing or decreasing either BO=7 or SO=7 by 1 resulting in 50% duty cycle and increasing or decreasing both BO=SO by 1 achieving a 100% duty cycle. Those scenarios examined for the three arrival rates applications applied on seven PANs which differ in number of nodes and each experiment are conducted five times. Network simulation parameters are presented in Table I.

TABLE I. QUALNET SIMULATION PARAMETERS FOR SEVEN PAN SCENARIOS

Parameter	Value
Physical and MAC model	IEEE 802.15.4
Area	80 m * 80 m
Number of nodes	2-8
Transmission range	10 m
Simulation time	1000 s
Channel Frequency	2.4 GHz
Data rate	250 kbps
Energy model	MICAZ
Antenna Height	0.08
Traffic	CBR
Payload size	50 byte
Arrival Rate	0.1,1, 2 seconds
BO and SO values	(6,6) to (8,8)

IV. SIMULATION RESULTS

The following subsections discuss the results obtained for each performance metric. We depict the results for four PAN scenarios due to space consideration.

A. *Total Energy Consumption (mWh)*

Four factors contribute in energy consumption and all need to be considered in investigating total energy consumption behaviour. Those contributors are the total energy a node consumes in packets transmission and reception, and the energy it dissipates during its idle listening and its sleep mode. Figures 1 to 4 depict energy consumption behavior for the four scenarios starting from one client up to seven clients.

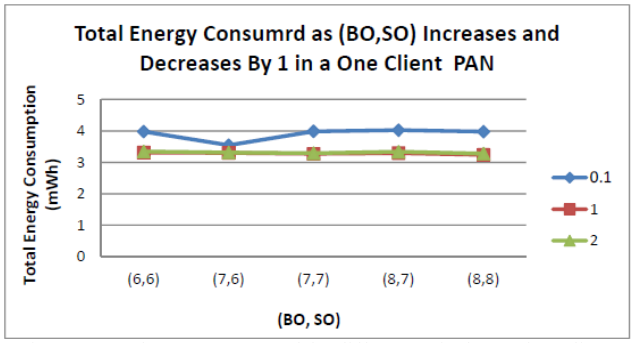


Figure 1. Total energy consumed for different arrival rates for 1client.

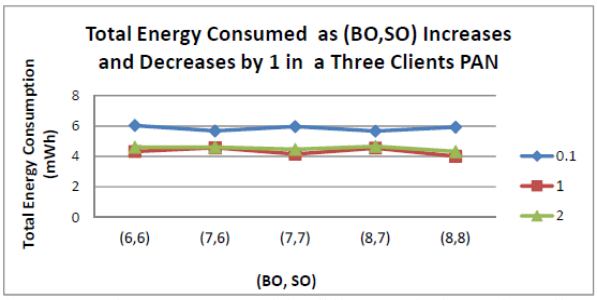


Figure 2. Total energy consumed for different arrival rates for 3 clients.

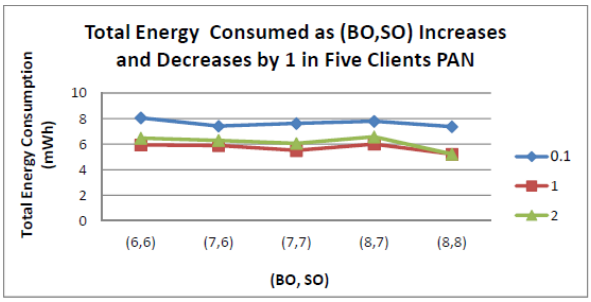


Figure 3. Total energy consumed for different arrival rates for 5 clients.

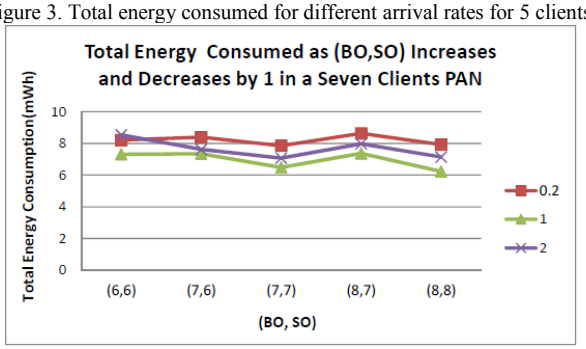


Figure 4. Total energy consumed for different arrival rates for 7 clients.

Very high arrival rate means frequent transmissions; hence, the dominating factor in energy consumption is the transmission power. However, after 500s, if the coordinator decides to increase or decrease either BO, or SO by 1, then, the node has a chance to rest and sleep for half of the BI period; thus, node saves energy. On the other hand, in low arrival rates applications, there would be more of a chance for free time while node is idle, which results in idle listening, hence, the dominating power consumption factor is idle listening. However, after 500s, if the coordinator decides to increase or decrease either BO, or SO by 1, then, node shall sleep for half of the BI period, and defer the transmission process for the next CAP if it receives a packet at the end of CAP. As the number of nodes increases, all the transmitting nodes may collide, which will cause retransmissions and as a result, energy consumption will increase. The obtained results show that for all arrival rates, increasing both BO and SO by 1 after 500s decreases energy consumption in comparison to the case of decreasing both BO and SO; this is because BI becomes longer, which in its turn, decreases beacon overhead at PANc.

B. *Average End-to-End Delay (s)*

Figures 5 to 8 depict average end-to-end delay behavior for the four scenarios starting from one client up to seven clients.

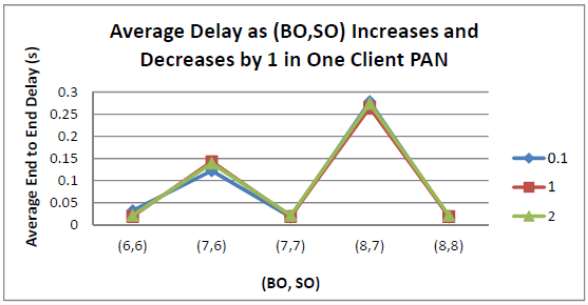


Figure 5. Average end-to-end delay for different arrival rates for 1 client.

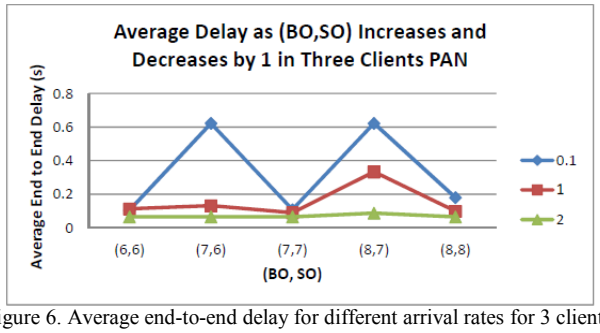


Figure 6. Average end-to-end delay for different arrival rates for 3 clients.

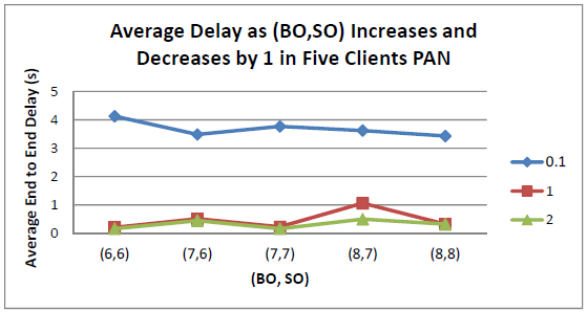


Figure 7. Average end-to-end delay for different arrival rates for 5 clients

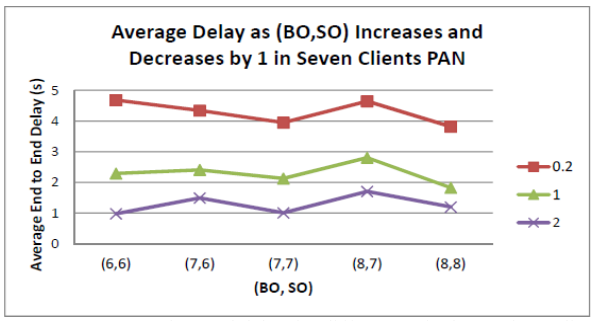


Figure 8. Average end-to-end delay for different arrival rates for 7 clients.

Despite the arrival rate, after 500s, if either BO or SO values decrease by 1, then, the duty cycle decreases to half, which cause nodes that have no time to complete their work in the current CAP, to wait for the next active period. This gets worse in active networks. As arrival rate increases, delay is less affected if the duty cycle decreases. However, if both BO and SO values are increased or decreased by 1, this achieves a full percent duty cycle, which means that nodes have more opportunity to use the medium, and this will produce the second delay factor, that is, the backoff delay. Backoff delay increases as the number of nodes increases. As the number of nodes increases, the possibility of collision increases, especially in short duty cycles, and hence, throughput decreases. This can be noticed at very low arrival rate values, such as 0.1 s. This explains why in PANs with 5 and 7 clients, delay decreases for the 0.1 s arrival rate application; it is simply because the number of the successfully arrived packets at the coordinator in a second decreases dramatically in active networks

C. *Throughput (bit/s)*

Figures 9 to 12 depict throughputs' results for the four scenarios starting from one client up to seven clients.

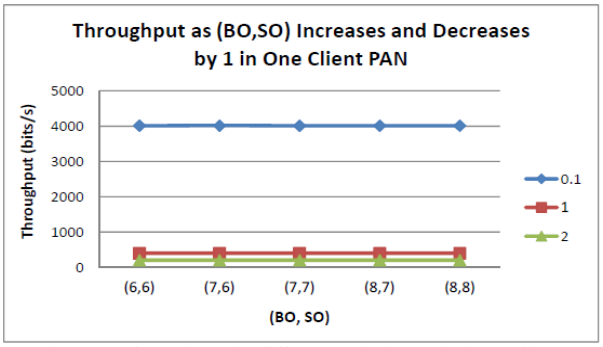


Figure 9. Throughput for different arrival rates in for 1 client.

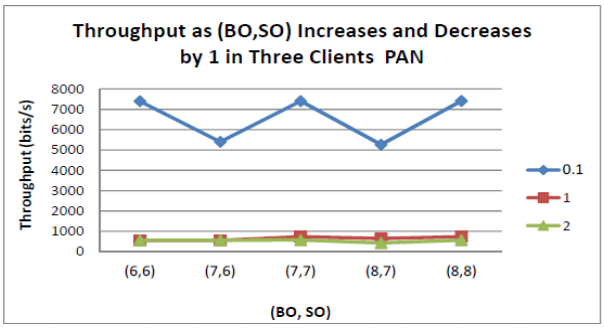


Figure 10. Throughput for different arrival rates for 3 clients.

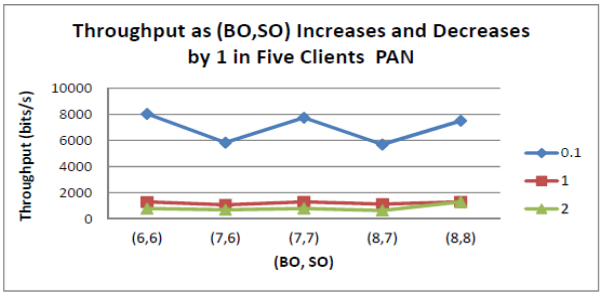


Figure 11. Throughput for different arrival rates for 5 clients.

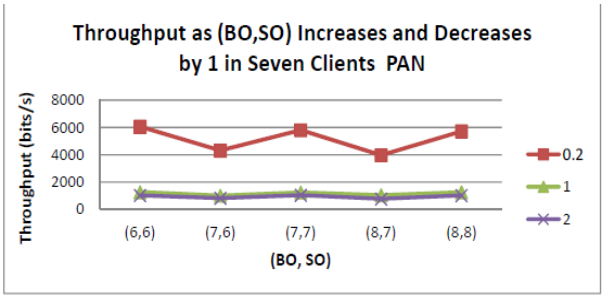


Figure 12. Throughput for different arrival rates for 7 clients.

Increasing or decreasing either BO or SO by 1, lessens the duty cycle, which cause nodes to wait for the next CAP

in order to accomplish the uncompleted transmissions. This causes severe collision at the beginning of the next superframe structure, which increases as the number of nodes increases. Obviously, such situation adversely affects throughput. This case is obvious in very active networks that follows 0.1s arrival rate. It is simply because more packets will be lost.

Increasing both BO and SO by 1 offers nodes longer CAP. This increases throughput, which is noticed as the number of nodes increases. As the duty cycle decreases, both delay and throughput are badly affected. This result is two folded: first, if we want to improve the performance, we need to avoid low duty cycles by having both BO and SO values close to each others as much as possible. In other words, we have to increase the duty cycle. Second, enhancing delay behavior through increasing the duty cycle, shall improve throughput behavior accordingly. This means that, we can propose an algorithm that enhances throughput just by enhancing delay or vice versa, this can be achieved by increasing the duty cycle.

V. CONCLUSION AND FUTURE WORK

WSN applications are diverse and can be classified according to the percentage of the duty cycle they work through and according to their packets arrival rate. Most of WSN applications follow IEEE802.15.4 standard. Performance of any WSN application is basically affected by the chosen (BO, SO) combination, thus, in order to achieve optimal performance, different types of WSN applications have to be aware of which (BO, SO) combination to activate. Decreasing the duty cycle is beneficial for active networks as the sleep periods offer the nodes the opportunity to rest; this is obvious in WSNs applications that follow 0.1 s. However, despite that decreasing duty cycle affects applications total energy differently, both delay and throughput behaviors are consistent for all arrival rates. In other words, as the duty cycle decreases by fixing BO while decreasing SO or vice versa, both metrics are adversely affected. As a future work, we suggest to design a general algorithm that adaptively improves the three metrics as the duty cycle increases irrespective of application arrival rate. Moreover, we need to investigate low and high arrival rates WSNs applications behavior in order to select both the optimal and the cut-off (BO, SO) combinations for each application. Such results are needed to develop an adaptive and general algorithm that adjust the (BO, SO) combination in accordance with the changes in the arrival rate.

REFERENCES

[1] IF. Akyildiz, W. Su, Y. Sankarasubramaniam, and A. Cayirci; A survey on sensor networks, Communications Magazine 2002. Atlanta, GA, USA, vol. 40(8), pp. 102-114, 2002.

[2] L. Selavo, A. Wood, O. Cao, T. Sookoor, H. Liu, A. Srinivasan, and J. Porter," wireless sensor network for environmental research", Proc. the 5th international conference on Embedded networked sensor systems. Sydney, Australia Nov. 2007, pp. 103-116.

[3] A. Koubaa," Promoting Quality of Service in Wireless Sensor Networks", (Submitted for receiving Habilitation Qualification in Computer Science) National School of Engineering, Sfax, Tunisia, 2011.

[4] SC. Ergen, "ZigBee/IEEE 802.15. 4 (Summary)", [accessed April 2013], Available from pages.cs.wisc.edu/~suman/courses/838/papers/zigbee.pdf.

[5] P. Park, C. Fischione, and K. H. Johansson, "Adaptive IEEE 802.15. 4 protocol for energy efficient, reliable and timely communications", Proc. the 9th ACM/IEEE international conference on information processing in sensor networks. Stockholm , April 2010, pp. 327-338.

[6] F. Shu, T. Sakurai, HL. Vu, and M. Zukerman, "Optimizing the IEEE 802.15. 4 MAC", Proc. IEEE Region 10 Conference (TENCON). Hong Kong,Nov. 2006, pp. 1-4.

[7] P. Patro, M. Raina, V. Ganapathy, M. Shamaiah, and C. Theiaswi , "Analysis and improvement of contention access protocol in IEEE 802.15. 4 star network", Proc. Mobile Adhoc and Sensor Systems (MASS 07). IEEE Internatonal Conference. Piza, Italy, Oct. 2007, pp. 1-8.

[8] H. Deng, J. Shen, B. Zhang, J. Zheng, J. Ma, and H. Liu, "Performance Analysis for Optimal Hybrid Medium Access Control in Wireless Sensor Networks". Proc. Global Telecommunications Conference (GLOBECOM 08). LA, USA, Nov. 2008, pp. 1-5.

[9] E. Casilari and J.M. Cano-Garcí, "Impact of the Parameterization of IEEE 802.15. 4 Medium Access Layer on the Consumption of ZigBee Sensor Motes". Proc. The Fourth International Conference on Mobile Ubiquitous Computing Systems, Services and Technologies (UBICOMM 2010). Florence, Italy, Oct. 2010, pp. 117-123.

[10] X. Li, CJ. Bleakley, and W. Bober, "Enhanced Beacon-Enabled Mode for improved IEEE 802.15. 4 low data rate performance", Wireless Networks 2012, vol. 18, pp. 59-74.

[11] M. Salayma, W. Mardini, Y. Khamayseh, and M. Yassein, "Optimal Beacon and Superframe Orders in WSNs". Proc. The Fifth International Conference on Future Computational Technologies and Applications (IARIA 2013), FUTURE COMPUTING 2013, May 2013, Valencia, Spain, pp. 49-55.

[12] M. Neugebauer, J. Plonnigs, and K. Kabitzsch, "A new beacon order adaptation algorithm for IEEE 802.15. 4 networks", Proc. The Second European Workshop on Wireless Sensor Networks. Belgium 2005, pp. 302-311.

[13] A. Koubaa and M. Alves, E. Tovar, "A comprehensive simulation study of slotted CSMA/CA for IEEE 802.15. 4 wireless sensor networks", IEEE WFCS, 2006, pp.63-70.

[14] F. Charfi, and M. Bouyahi, "Performance evaluation of beacon enabled IEEE 802.15.4 under NS2", arXiv preprint arXiv 2012, pp. 1204.1495.

[15] SA. Khan and FA. Khan, "Performance analysis of a zigbee beacon enabled cluster tree network", Proc. Third International Conference on Electrical Engineering (ICEE'09). LahoreApril 2009, pp. 1-6..

[16] J. Hoffert, K. Klues, and O. Oriih "Configuring the IEEE 802.15. 4 MAC Layer for Single-sink Wireless Sensor", Washington University in St. Louis, 2005.

 ISBN: 978-1-61208-296-7

Challenges in Securing Wireless Sensor Networks

Heshem A. El Zouka
Department of Computer Engineering, College of Engineering and Technology
Arab Academy for Science & Technology and Maritime Transport,
Alexandria, Egypt
helzouka@aast.edu

Abstract — **Wireless sensor networks are highly prone to faults and errors due to their restrictive capabilities and limited resources. In this paper, some of the challenges facing the wireless sensor networks such as security, routing, computing capability, and battery power, will be discussed. We will focus on both security and authentication protocols in attempting to provide a security solution against known malicious attacks. Furthermore, utilizing the existing security protocols in wireless sensor networks has led us to propose a framework that incorporates authentication and encryption techniques in accordance with mobile agent technology in a way that reduces the energy consumption in such low power wireless sensor networks.**

Keywords-Security; Authentication; Routing; Energy efficient; Energy conserving; Agent-based architecture.

I. INTRODUCTION

Wireless Sensor Networks (WSNs) have recently attracted great attention due to their wide use in modern applications, such as military radio applications, environmental monitoring, and heath care industry. WSNs may consist of a large number of tiny and inexpensive sensor nodes that are densely deployed where the neighboring nodes may be very close to each other in a large area, collecting salient information from the target field. In a WSN, the base (sink) node is carefully positioned, so that it can communicate with all network nodes using the proper methods and protocols [1].

In general, the sensor nodes share a common predefined purpose to sense a predefined surrounding area being monitored and the sensed data is aggregated and transmitted to the base station where analysis can be performed to obtain some summary measures of the information provided by the sensor nodes. Thus, security became a necessity since sensor nodes are being small and less capable to provide tamper-resistant and their deployment is left unattended once dispersed over unprotected areas.

It is, therefore, difficult to determine whether the node participating in the routing domain is Byzantine [2] or indeed an authenticated node. The security risks and power consumption remain high-priority issues in designing WSNs [3]. Despite of these two issues, WSN has many features including flexible deployment, self-organizing, high reliability, and low-cost.

In section two, we are going to present the trends and challenges facing the development of new security models in WSNs. The third section presents the proposed security models with their respective advantages and disadvantages in details. The fourth section gathers everything together; our implementation is discussed along with all the simulation results obtained and a comparison of the results is presented. Section six draws a conclusion and future works.

II. LITERATURE SURVEY

WSN and Ad hoc networks [4] are getting more and more popular now. In fact, WSNs are inherited from ad hoc networks in the sense of being wide spread over geographical locations.

Ad hoc networks have no predefined, fixed infrastructure; they have no centralized controller or a fixed router [5]. Unlike wired and Mobile Ad hoc Networks, WSNs are infrastructure-less and can work in any environment as compared to the traditional computer networks.

The main difference between ad hoc networks and WSNs is that a WSN has reduced capabilities of the sensing nodes, such as reduced energy, computational power, no mobility, and limited transmission range [6]. The main components of sensor nodes are shown in Figure 1.

A. Energy Conservation

The limitations and the specific architecture of sensor nodes call for energy efficiency and secure authentication protocols. The feasibility of these inexpensive sensor networks is driven by the advances in Micro Electromechanical Systems (MEMS) technology [7], combined with low power, low cost CMOS logic. Through sensor networks, we are able to analyze data more accurately, gain more knowledge and improve our awareness of risk management issues.

Hence, it is quite essential to secure such data and employ strong authentication models while maintaining a reasonable level of energy dissipation.

 ISBN: 978-1-61208-296-7

In order to guarantee proper operation, sensor nodes must be supported by a power unit, which is usually in the form of a battery, however, in some cases, energy can be regenerated using solar cells. Energy from power scavenging techniques may only be stored in secondary (rechargeable) batteries and this can be a powerful combination in wireless sensor node environments where standard maintenance procedures like battery changing are impractical. Furthermore, when energy saving technique such as Dynamic Voltage Scaling (DVS) is used, it preserves energy and power.

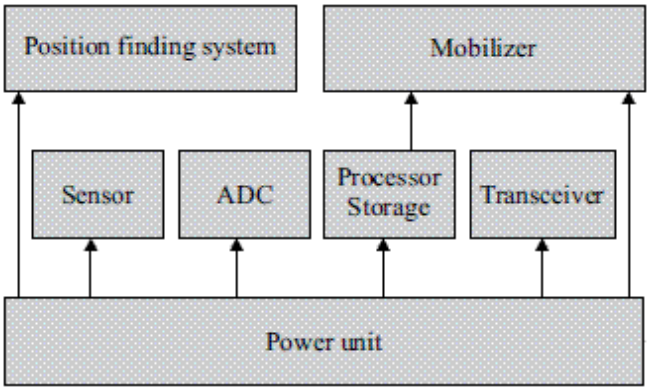


Figure 1. Sensor node components

During normal operation, every node periodically wakes up from sleep mode to communicate with its neighbors, and then goes to sleep again until the beginning of the next frame. In addition, due to the limited energy supply of the sensor nodes, it is assumed to be limited to use Public Key Cryptography (PKC) [8] in WSNs.

Recent research shows that PKC consumes a significant amount of computing resources and power. On the other hand, many studies have proved the falsehood of such assumptions and the ability to form a more secure communication under certain arrangements [9].

Meanwhile, new messages are temporarily queued to allow in-transit messages from node to node. Sensor nodes communicate with each other using a Request To Send / Clear to Send (RTS/CTS), and Acknowledgement (ACK) scheme, which provides both collision avoidance and reliable transmission scheme [10].

B. P2P Vulnerabilities

A Peer-to-Peer (P2P) network is a special type of network that has many characteristics over client-server manner. All wireless nodes are considered clients and servers and communicate directly with each other without an intermediate centralized hub. They all provide and consume network services. Peers are also routers as data passes through them to reach the destination node.

The difference between P2P networks and Cluster-based networks is illustrated in Figure 2 (a) and (b). P2P in WSNs has emerged to realize a computing architecture with no single-point of failure [11].

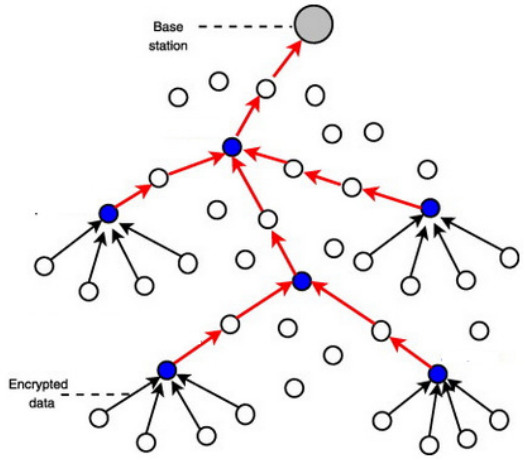


Figure 2. (a): Peer-to-Peer Communications in WSN

Its inherent scalability and lack of administration has made its cost virtually low. However, these features introduced a new set of vulnerabilities, such as false data injections and false data filtering, which have become an important issue for all WSNs. As opposed to the client-server networks, P2P has no always-on infrastructure servers [12].

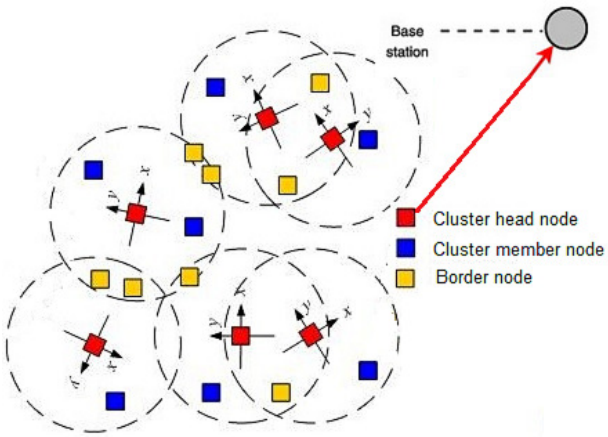


Figure 2. (b): Cluster Based Communications in WSN

The decentralized data source architecture eliminates the single-point of failure. This allows nodes to enter and leave the network frequently. Hence, a P2P network acts as a dynamic architecture. Moreover, decentralized networks have security issues since each node is responsible for controlling their data and resources. Hence, unauthenticated node within the cluster may send data that turns out to be a virus. This paper presents a combination of P2P and WSN networks to ease the development of models that rely on WSN functionality. As a result, we propose a network of, say, sensors, or P2P nodes, which allow integrating widely, distributed sensor nodes and other computing devices.

 ISBN: 978-1-61208-296-7

III. SECURITY IMPLICATIONS AND SOLUTIONS

In this paper, we performed some initial tests on the effectiveness of the envelope model that aims at securing the communication using Asymmetric Public key model, such as RSA algorithm, among the upper hierarchy, and symmetric model cryptography using AES among nodes. In this model, the network is divided into clusters and all the nodes inside each cluster will act as Cluster Heads (CHs) one with a different number of cycles in a fashion that maintain energy balancing. All nodes that are not CHs only communicate with their CH node using AES encryption model. On the other hand, all CHs are communicating with each other using strong encryption models, such as RSA cryptosystem, that make it essentially impossible for third party to decrypt. Therefore, the proposed security model guarantees data confidentiality as the keys cannot be detected due to the use of PKI cryptosystem. No malicious node can intercept or read the encrypted data of the sensed node. In later section of this paper, we will show how the performance of the sensor network is analyzed in terms of security, energy efficiency, and packet delivery rate.

We implemented our models based on the following assumptions: (1) the sensor nodes are deployed randomly with uniform distribution and (2) All nodes are homogeneneous and have the same initial energy. The energy model we have used is similar to [13], where E_{elec} = 50 nJ/bit as the dissipation of energy is used to run the transceiver circuit, and amplification of E_{elec} = 100 pJ/bit/m^2. The expansion of energy during transmission and reception of a k bit message is given by the following two equations; where, λ represents the path loss exponent and has a value ≥ 2.

$$E_{trans}(k, d) = E_{elec} * k + \in_{amp} * k * d^{\lambda} \quad (1)$$

$$E_{rec}(k) = E_{elec} * k \quad (2)$$

A. Envelope Routing Protocol

The routing attacks that affect the WSNs are also considered when implementing the envelope model. In case an attacker node changes any of the fields in a message or the destination of the message, such an attack is detected as the MAC algorithm performs the integrity check procedures. Another possible attack is sending malicious messages to nodes of the cluster. In this model, all nodes communication must go through the CH/BS, which keeps a list of all authenticated nodes. Therefore, the source of any message is checked against that list, which prohibits attackers from sending false data.

In our proposed model, we decided to employ the LEACH routing protocol. LEACH is currently one of the most famous communication protocols for WSNs [14]. LEACH is self-organizing and adaptive clustering protocol that is characterized by its feature of distribution the energy load among the sensing nodes within each cluster. Clusters, then, are created dynamically through a randomized rotation of cluster heads after each round. Data is aggregated by the CH and is sent to the BS after each round. LEACH assumes that the BS is fixed and located far from the sensors and all nodes in the network are homogeneous and energy constrained.

A cluster hierarchy will also aid us in utilizing an elected cluster head CH to perform asymmetric cryptosystem over the communication with other cluster heads, which decreases the communication overhead and bounds the asymmetric encryption operations to the elected cluster heads only. On the other hand, all nodes within the same cluster communicate using symmetric cryptosystem.

In case the attacker sends malicious data to cluster nodes, it will be detected since all the communicating nodes are trusted and authenticated by the base station itself as mentioned before. Hence, the source of any message is checked against that list, which prohibits attackers from sending malicious packets using forged or spoofed source addresses.

B. Envelope Authentication Protocol

In this section, we will show how the Mobile Agent (MA) is employed to enhance the authentication protocol of the sensing nodes. The design is similar to [15] in that the authentication of sensor nodes is controlled by the mobile agents platform using public key infrastructure and digital signature. The intelligence of a MA can be used to make dynamic decisions such as optimizing the travel plan, finding the next destination, and detecting link failures [16]. One of the major challenges of the deployment of MAs in WSNs includes secure transmission of agent as well as preventing unauthorized access to resources between communicating sensor nodes. However, MAs that move around the wireless networks are not safe because the remote hosts that accommodate the agents initiate all kinds of attacks.

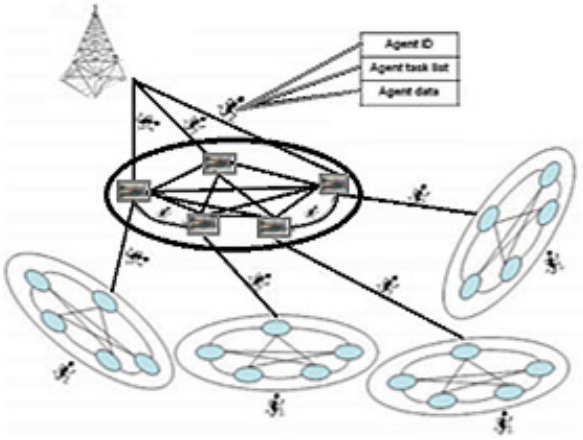


Figure 3. Middleware Mobile Agent Based Architecture

 ISBN: 978-1-61208-296-7

In addition, dispatching a set of low-level software modules to the sensor nodes is not an easy task, since it needs to interact with the sensor hardware devices within the network. Thus, to simulate this architecture, it is assumed that the modules are hard-coded into the memory of each node. Although some architecture allow the software developers to employ a node level OS, the developer still has to perform a single executable image to be configured manually into each node [17]. Overall, there is a strong need for developing a middleware environment that supports dynamic programming and simplify the complicated authentication / routing protocols in the sensor networks. Challenges in the development of Middleware approaches for WSNs constitute an emerging topic that is being investigated by a number of researchers [18].

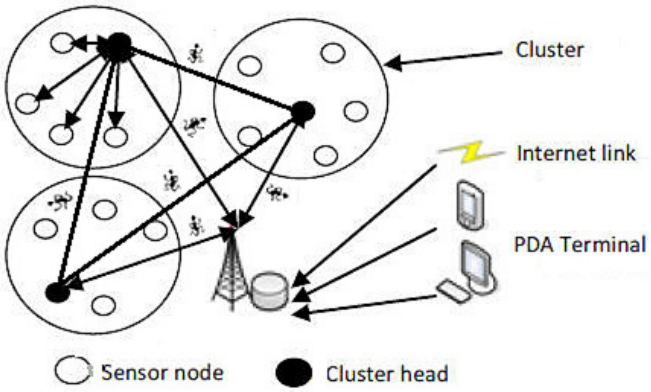


Figure 4. Envelope MA Based Architecture

Many of these approaches are based on mobile code (byte code) that runs on top of TinyOS [19], where a lightweight operating system specifically designed for such low power WSNs.

Application programs are broken up into small instruction segments, each of a single byte long, making it easier to be replicated and propagate through the network. Sending short instruction segments allows rapid deployment of aggregated message authentication code, as shown in Figure 3. Propagation of code segment is done whether by comparing code versions between neighbors (whenever possible) to update the routing information or by pursuing dynamic programming of sensor nodes. However, node's ability to allow code motion is highly limited.

To overcome this problem, our implementation utilizes how MA is used to perform the authentication mechanism without risking the security of the agents. As shown in Figure 4, the suggested architecture has been offered in two layers that aim to secure the execution results of MAs. The first layer aims to preserve the authenticity of the node via a combinational distributed and centralized agent server among the CHs hierarchies.

On the other hand, the second layer utilizes the authenticity of the agents on all nodes of the cluster. In this way, multiple MAs could be created and dispatched to the sensor nodes as soon as they are requested. For example, the central station could dispatch mobile software agents to perform authentication test and check communication between the sensor nodes. This will lead to minimize the communication overhead by eliminating unnecessary communications.

Data analysis and fault diagnosis MA algorithms can migrate over the network to perform collaborative monitoring and computational tasks. As shown in Figure 4, there is a centralized agent server per each cluster of the network, which maintains the security of the traversing agents within cluster.

That makes it essentially impossible for third party to attack the agents. In addition, the model guarantees data confidentiality as the keys cannot be detected due to the applied PKI in the clustering protocol, as discussed in the previous section of this paper.

C. Applied Compression and Hashing Algorithms

Due to energy limitation, a WSN faces to the challenge of lifetime. In order to minimize the energy consumed in the data communication, we need to decrease the time elapsed in data transmission.

In this section, we will investigate whether it is possible to save energy and load balancing in the WSNs by applying a compression technique over the data before sending. Thus, minimizing the size of the data, so that the radio transmission time will be reduced and the message will be less likely to be lost [20]. Compressing data helped us in saving energy as the data size to be communicated is smaller. Since accessing memory is quite expensive in energy, we searched for a compression algorithm that provides less memory access during execution time.

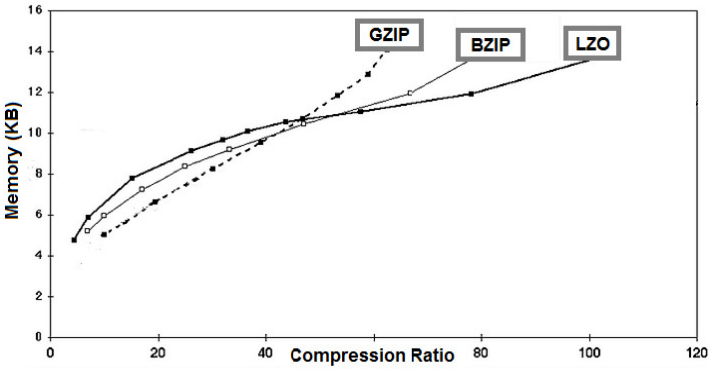


Figure 5. Speed and Compression Ratio Comparison

Figure 7 shows a comparison between different data compression algorithms, in terms of compression ratio and percentage root mean square difference.

Gzip is short for GNU zip [21]. It was created by Jean-Loup Gailly and Mark Adler. It is based on the Deflate algorithm and it usually used to find duplicate strings in the input data. The second occurrence of compression, a string

is replaced by a pointer that points to the previous string in the form of distance and length pair. Literals or match lengths are compressed with one Huffman tree [22], and match distances are compressed with another tree.

The trees are stored in a compact form at the start of each block. Duplicate strings are found using a hash table. All input strings of length 3 are inserted in the hash table. A hash index is computed for the next 3 bytes. If the hash chain for this index is not empty, all strings in the model will be compared with the current input string, and the longest match is selected. In this model, gzip for data compression was used as it offers fast compression technique, and executes fewer instructions per bit.

IV. COMPARATIVE PERFORMANCE ANALYSIS

Traditionally, the three main techniques for analyzing the performance of wired or wireless networks are analytical methods, computer simulation, and physical measurement. However, because of many constraints imposed on sensor networks, such as energy limitation, algorithms for sensor networks tend to be quite complex and usually defy analytical methods that have been proved to be fairly effective for traditional networks.

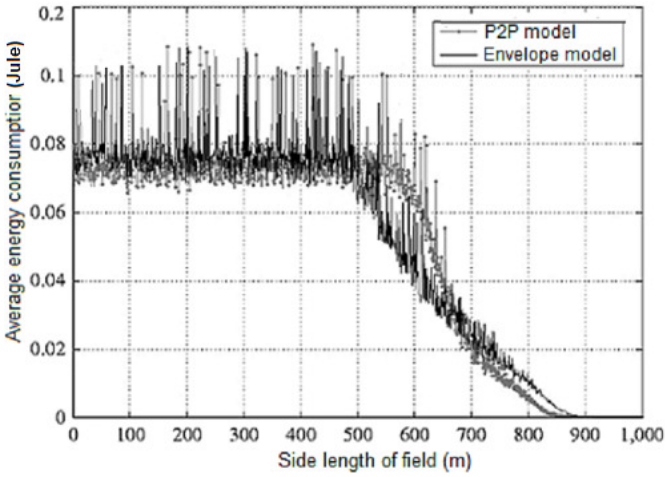


Figure 6. Total Energy Consumption

Furthermore, few sensor networks have come into existence, for there are still many unsolved research problems, so measurement is almost impossible. It appears that simulation is the only feasible approach to the quantitative analysis of sensor networks. In order to compare the implemented routing approaches; a powerful simulator was needed to measure the energy dissipation, end-to-end delay and throughput of both protocols.

The mixed mode in NS-2 using OTcl language [23] with underlying C++ classes made it complex to be used in our protocols. OMNET provided a simpler yet robust options, it enabled us to build the communication layers of the sensor nodes in an ease approach [24].

Figure 6 shows the energy consumption of a network when simulating the proposed envelope encryption model.

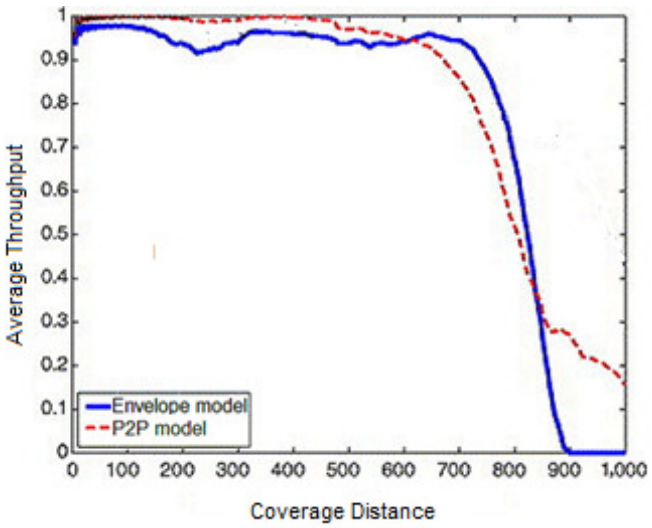


Figure 7. Average Throughput as a Function of the Field Size

The energy of the envelope model is the sum of energy of all nodes including the energy of the elected cluster heads. The consumed energy is shown to be lower for the envelope model as compared to the other models which use a public key cryptographic algorithm. In addition, the analysis provides a clear view of the percentage of nodes, which have considerable amount of energy compared to cluster head nodes of the nearly consumed ones. In the envelope model, the energy is more distributed over all the nodes, and more than 70% of the nodes in the envelope model preserved their energy as the energy is consumed mostly around the CHs. On the other hand, the model shows a more balanced energy consumption model, since all the nodes perform their own data transmission.

The radio of each non cluster head node is turned off till the node allocated time is reached, thus minimizing energy dissipation in these nodes. Once all data is received from all expected nodes in the cluster, the CH aggregates these data using the proposed MA mechanism and forwards them to the BS as one message in an effort to reduce the communication overhead. The storage requirement in the envelope model is also high, especially for CHs, since CHs store all the symmetric keys of the cluster nodes as well as any data to be sent or compressed. Each node in the node-to-node (n2n) connection model sends one message and as a result, asymmetric encryption is limited and this requires higher storage and computing complexity. Figure 7 illustrates the average throughput measured at the BS, the average throughput of the envelope model appears to be higher than the n2n connection model due to the aggregation of all data at the CHs. The envelope model provided a slight increase of 4% in the average throughput. The

 ISBN: 978-1-61208-296-7

analysis also showed that the average introduced delay in n2n approach is 27.4 sec while in the envelope model it reaches 34.5 sec due to the proposed security protocols.

V. CONCLUSION AND FUTURE WORK

In this paper, we surveyed the challenges and the proposed solutions and approaches for the security and routing protocols of WSNs. We then proposed a framework that secures the communications between the wireless nodes. In the first experiment, we implemented an envelope model that uses RSA asymmetric algorithm along with a symmetric algorithm in a hybrid manner to preserve the energy and increase the life time of WSN. In the second experiment, we examined how MA is employed to enhance the authentication protocol of the sensing nodes. To improve the performance of a middleware environment, the proposed envelope architecture is implemented using two security layers, one for establishing authenticity and one for generic trust that authenticates the distribution agents.

The analysis showed that the proposed envelope protocol provides a significant increase in the life of the entire network as more than 70% of the nodes reserved their energy while the consumption is limited to the CHs. Evaluating the effectiveness of employing a strong compression technique, the analysis showed that the GZIP experience fewer losses by sending fewer instructions per packet and the resulting compressed data size was roughly 2 times less than of original instruction data set size.

In the future, we plan to increase the scale of the network and to use more than one base station in our network, also we plan to make our protocols aware of data freshness by adding time stamp to the authenticated packet. We will explore more in dynamic compression techniques as well. Additionally, we plan to study the performance of our model on different motes and build a comparison over different architectures.

VI. REFERENCES

[1] A. Dearle, D. Balasubramaniam, J. Lewis, and R. Morrison, "A Component-Based Model and Language for Wireless Sensor Network Applications," 32nd Annual IEEE International on Computer Software and Applications, Turku, August 2008, pp. 1303-1308.

[2] P. Kiran Sree and I. Ramesh Babu, "Towards a Cellular Automata Based Network Intrusion Detection System with Power Level Metric in Wireless Adhoc Networks," Proceedings of 2008 International Conference on Advanced Computer Theory and Engineering, 2008, pp. 1071-1075.

[3] C. Chang, D. J. Nagel, and S. Muftic, "Assessment of Energy Consumption in Wireless Sensor Networks: A Case Study for Security Algorithms," In 4th IEEE International Conference on Mobile Ad Hoc and Sensor Systems (IEEE MASS 2007), Pisa, Italy, October 2007, pp. 1-6.

[4] T. El Maliki and J. M. Seigneur, "A security adaptation reference monitor (SARM) for highly dynamic wireless environments," in Proceedings of the 4th International Conference on Emerging Security Information, Systems and Technologies (SECURWARE '10), July 2010 , pp. 63–68.

[5] C. Sreedhar, S. Vema, and P. Kasiviswanath, "A Survey on Security issues in Wireless ad hoc Routing Protocols, International Journal 2(2), 2010, pp. 224-232.

[6] L. Badia, M. Conti, S,K. Das, L. Lenzini, and H. Skalli, "Routing, interface assignment and related cross-layer issues in multiradio wireless mesh networks," In Guide to wireless Mesh Networks, Springer, London, 2009, pp. 147-170.

[7] MicroEelectroMechanical Systems (MEMS): http://mems.sandia.gov/. [Retrieved on June, 2013].

[8] W. Diffie and M. E. Hellman, "Multiuser cryptographic techniques," In Proceedings of National Computer Conference, New York City, AFIPS, June 1976 , pp. 109-112.

[9] S. Othman, A. Trad, and H youssef, "Performance Evaluation of Encryption Algorithm for Wireless Sensor Networks," International Conference on Information Technology and e-Service (ICITeS), March 2012, pp. 23-35.

[10] A. Pandey and R. Tripathi, "A Survey on Wireless Sensor Networks Security," International Journal of Computer Applications, vol. 3, no. 2, June 2010, pp. 8887 - 8975 .

[11] J. Sen, A Trust Based Detection Algorithm of Selfish Packet Dropping Nodes in Peer to Peer Wireless Mesh Network, In Recent Trends in Network Security and Applications, Springer, New York, 2010, pp. 528-537.

[12] T. Kavitha and D. Sridharan, "Security Vulnerabilities in Wireless Sensor Network: A Survey," Journal of Information Assurance and Security, vol. 5, no. 1, July 2010, pp. 31-44.

[13] S. Mostafa, H. El Zouka, and M. Abouelnasr, "Hybrid Encryption Secure Routing Protocols for Wireless Sensor Networks," Proceeding of the ISCA, First International Conference on Sensor Networks and Applications (SNA), San Francisco, November 2009, pp. 109-114

[14] S. Kumar, M. Singh, and D. Singh, "Routing Protocols in Wireless Sensor Networks - A Survey," International Journal of Computer Science & Engineering Survey (IJCSES), vol. 1, no. 2, November 2010, pp. 570-580.

[15] S. Srivastava and et al, "A Survey on Mobile Agent based Intrusion Detection System," IJCA Proceedings on International Symposium on Devices MEMS, Intelligent Systems & Communication (ISDMISC), New York, October 2011, pp. 19-24.

[16] M. Chen, T. Kwon, Y. Yuan, Y. Choi, and V. Leung, "Mobile Agent Based Directed Diffusion in Wireless Sensor Networks," Journal on Advances in Signal Processing, vol. 1, April 2007, pp. 219-242.

[17] G. Fortino, A. Garro, S. Mascillaro, and W. Russo, "Specifying WSN Applications through Agents Based on Events and States," International Conference on Sensor Technologies and Applications, Valencia, October 2007, pp. 463-468.

[18] F. Y. Xiong and L. Bai, "Interoperable Wireless Sensor Network Model Using Multi-Agent-Based Middleware," International Symposium on Intelligent Signal Processing and Communication Systems, Chengdu, December 2010, pp. 1-4.

[19] W. Munawar, M. Alizai, O. Landsiedel, and K. Wehrle, "Dynamic TinyOS: Modular and Transparent Incremental Code-Updates for Sensor Networks," In proceedings of the IEEE International Conference on Communications (ICC 2010), pp. 1-6.

[20] T. Arici, B. Gedik, Y. Altunbasak, and L. Liu, "PINCO: a Pipelined in-Network Compression Scheme for Data Collection in Wireless Sensor Networks," In Proceedings of 12th International Conference on Computer Communications and Networks (ICCCN)," October 2003, pp. 539-544.

[21] Gzip Home page. http//www.gzip.org [Retrieved on June, 2013].

[22] S. Mohajer, P. Pakzad, and A. Kakhbod, "Tight Bounds on the Redundancy of Huffman Codes," IEEE Information Theory Workshop (ITW), Punta del Este, Uruguay, March 2006, pp. 131-135

[23] The Network Simulator-NS-2. http://www.isi.edu/nsnam/ns [Retrieved on June, 2013].

[24] The OMNeT++ Simulator. http://www.omnetpp.org [Retrieved on June, 2013].

 ISBN: 978-1-61208-296-7

The Cloning Attack Vulnerability in WSN Key Management Schemes

Othmane Nait Hamoud
Ecole Militaire Polytechnique
Algiers, Algeria
o_naithamoud@esi.dz

Tayeb Kenaza
Ecole Militaire Polytechnique
Algiers, Algeria
ken.tayeb@gmail.com

Nadia Nouali-Taboudjmat
CERIST
Algiers, Algeria
nnouali@cerist.dz

***Abstract*—The majority of existing key management schemes in wireless sensors networks suffer from the physically compromised of nodes. This vulnerability allows an adversary to reproduce clones from a compromised node and to inject them into several locations on the network to conduct other types of attacks. Moreover, the joining of new nodes to the network (for maintenance), which is an inevitable step to prolong its life or to correct voids, presents the best opportunity to conduct such an attack. In this paper, we review several key management schemes and we highlight their weakness regarding the cloning attack.**

Keywords—WSNs; key management scheme; cloning attack; maintenance of WSNs.

I. INTRODUCTION

Wireless Sensor Networks (WSNs) present an effective solution in many areas (military, environment, etc.) thanks to their low cost manufacturing, their multiple usages and their ad hoc property. However, their deployment in open spaces, their transmission medium and the lack of infrastructure with a robust security, expose them to serious vulnerabilities.

To protect WSNs, several solutions have been proposed in the literature [12], where the central element is the key management scheme. This latter can be classified regarding the used encryption technique as symmetric, asymmetric, or hybrid. Each of these schemes, try to find a compromise between cost and efficiency, where their common goal is to allow each pair of neighboring nodes to establish a symmetric cryptographic key. These keys are then used to ensure the confidentiality of the exchanged data between nodes. However, the random deployment of a huge number of sensors renders unpredictable the resulting network topology. Therefore, one can never predict for each sensor its direct neighbors in order to pre-load it with the adequate Pair-wise Keys (PK), which makes the design of secure key management solutions a challenging task. Another difficulty lies in the authentication mechanism. This can be achieved through asymmetric cryptography, which is known for its requirements in terms of resources consuming. Hence, the convergence of the vast majority of key management approaches to the symmetric cryptography, when authentication is generally based on the pre-loading of sensors with a primary key.

However, these mechanisms suffer from physical sensors compromise. This vulnerability allows an adversary to reproduce clones from a compromised node and inject them in several places in the network (known as cloning or replication attack). These clones will then present themselves as legitimate nodes to infiltrate the network and then conduct other types of attacks [13]. Moreover, the joining of new nodes in the network (for maintenance) is an inevitable step to prolong its life and repair voids (not covered areas) or isolation of subnets that can arise both immediately after the initialization of the network or during its operational phase, presents the best opportunity to lead the cloning attack.

The remainder of this paper is organized as fellows. We introduce in Section 2 the cloning attack vulnerability in distributed key management schemes. In Section 3 we review some of these schemes where we highlight their weakness regarding the cloning attack. Section 4 concludes this paper.

II. CLONING ATTACK IN WSNS

To minimize the impact of the physical compromise of sensors in primary key based schemes, three alternatives have been proposed:

1. The primary key must depend on the geographic location of new nodes (localization based schemes) [5], [11]. In this case, the compromise of this key will not allow an adversary to inject its clones in locations other than those in which he compromised them.
2. The primary key must be checked by a trusted third party before joining a new node (centralized schemes) [6] [10].
3. The primary key must depend on time (time-based schemes) and/or on the use of one-way hash function [1], [2], [3], [4].

Generally, in localization based schemes, deployment is done in such way that each sensor has the coordinates of its location (known previously or given by a Global Positioning System (GPS)) or has its relative position regarding its neighbors in the case of deterministic deployment (e.g., grid deployment [7]). Therefore, information about the position of a sensor can be involved into the process of key establishment, which prevents from (or greatly limits) cloning attack in this category. In centralized schemes, the centralized control of maintenance operations can significantly reduce the impact of the cloning attack. However, these two categories present, respectively, several constraints regarding the deployment and the scaling.

Note that we are particularly interested in the third category (time-based schemes), since their schemes are the most cited in the literature. These schemes offer unlimited scalability, a relatively secure connectivity, and a decentralized management.

Unfortunately, these schemes are still subject to the cloning attack, where the impact is very important regarding

 ISBN: 978-1-61208-296-7

to the effort made by an attacker to conduct such attack. Indeed, to infiltrate a WSN and to carry out more attacks, we need just to compromise a single sensor, extract the secret information from its memory and reproduce several clones, and finally inject them in several locations of the network [13]. These clones will be able to legitimately communicate with old nodes as new nodes, and to conduct the attack that we call "direct cloning" or, wait for the deployment of new nodes to present themselves to the latter as old nodes, and thus lead the attack that we call "indirect cloning" (Fig. 1).

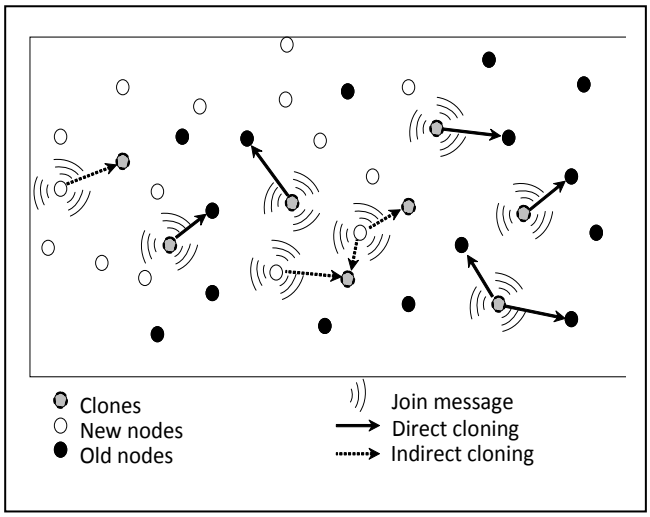


Figure 1. Direct and indirect cloning attack.

Time-based schemes, although they claim resisting to the cloning attack, they finally resist only to the direct cloning attack. Note that, to our knowledge, the indirect cloning has never been discussed in the literature.

III. REVIEW OF SOME EXISTING TIME-BASED SCHEMES

In this section, we will review some time-based key management schemes, where we particularly highlight their weakness regarding the cloning attack and its impact on the security of WSNs.

A. LEAP

Localized Encryption and Authentication Protocol (LEAP) [1] is one of the most referenced protocols in the literature. It presents a comprehensive key management scheme based on a primary key, under the assumption that this key will be erased after the network initialization. It has been designed to support multiple communication modes: unicast, broadcast, local and global broadcast.

In order to bypass the use of the BS in each operation of key establishment or authentication between nodes, authors of LEAP have based their scheme on a primary key K_{IN}, which is pre-loaded in nodes.

They assume that there is a time limit T_{min}, needed to compromise a node, and another time limit T_{est} required by a newly deployed sensor to discover its immediate neighbors, to establish PK with each of them and to erase K_{IN}. Zhu et al. [1] have proposed two variants of this scheme:

- Basic scheme

In this scheme, it is assumed that an attacker cannot recover K_{IN} before T_{min} (i.e., $T_{est} < T_{min}$). Before deployment, the BS pre-loads the primary key K_{IN} on a node u, which is then used to derive a primary key $K_u = f_{K_{IN}}(u)$ where f is a pseudo-random function. Once deployed, the node u broadcasts a message containing its identity u and initiates a timer, which will be triggered when T_{min} expires. The response of a node v includes its identity and the MAC (calculated using K_v) of u and v:

$$u \rightarrow *: u \quad (1)$$

$$v \rightarrow u: v \mid MAC(K_v, u|v) \quad (2)$$

The response of v can be only authenticated with the primary key K_v, which is derived as follows: $K_v = f_{K_{IN}}(v)$. As the node u has K_{IN}, it can also derive and verify the identity of v. Thus, the PK will be calculated as follows: $K_{uv} = f_{K_v}(u)$. Once T_{min} is expired, the node u erases K_{IN} and all primary keys of its neighbors, while it keeps its K_u.

- Extended scheme

In this scheme, authors assume that an attacker can recover K_{IN} before T_{est} (i.e., $T_{est} > T_{min}$). The basic idea in this case is to remove the dependency on K_{IN}, which will be replaced by a series of primary keys corresponding to time intervals (T_1, $T2$, ..., T_M) of new nodes joining. Before deployment, the BS pre-loads a node u to be deployed in the interval Ti with the primary key K_{IN}^i, through which it derives its primary key $K_u^i = f_{K_{IN}^i}(u)$. In addition, the node u is pre-loaded with primary keys $K_u^j = f_{K_{IN}^j}(u)$ for all $i < j \leq M$. Once deployed, the node u starts the neighbor discovery and the establishment of PK. In the neighbor discovery, a node u begins by sending a message containing its identity u and the current interval i:

$$u \rightarrow *: u \mid i \quad (3)$$

As response, a neighbor v sends an acknowledgment Ack authenticated with the master key K_v^i of the current interval i:

$$v \rightarrow u: v \mid MAC(K_v^i, u|v) \quad (4)$$

Since u has K_{IN}^i, it can calculate K_v^i and check the Ack. Now, nodes u and v can calculate their PK : $K_{uv} = f_{K_v^i}(u)$. Once T_{min} is expired, the node u erases K_{IN}^i and all primary keys of its neighbors, while it keeps its K_u^i and primary keys of future intervals j.

Discussion

LEAP authors address the cloning attack only in the case when a new node compromise would be made before T_{min}, while the cloning attack can be done even if the compromise was made after T_{min}. In the basic scheme, the maintenance operation exposes K_{IN} to compromise, which may annul the

 ISBN: 978-1-61208-296-7

security of the entire network. Moreover, even if K_{IN} is not disclosed, the indirect cloning attack is possible against new nodes. Indeed, the primary key of a compromised node can be directly used by clones to authenticate theirs responses (2) to messages sent by new nodes during the neighboring discovery (1).

In the extended scheme, if the attacker compromises a node u after T_{est}, the primary key K_{IN}^{i} of a node u and its neighbors primary keys K_v^i are deleted, which prevents from the direct cloning attack. However, the attacker can recover primary keys K_u^j that correspond to future intervals j. Thus, he can clone the node u, and all copies will be able to present themselves as legitimate old nodes (4) to new nodes to be deployed in future intervals (indirect cloning).

For example, a node u is deployed at T_2, it is pre-loaded with K_{IN}^2 (Through which it derives its primary key K_u^2), and M-2 master keys K_u^3 ,…, K_u^M, corresponding to intervals T3, ..., Tm. After T_{min}, the node u do not delete master keys K_u^2 , K_u^3 , …, K_u^M in order to authenticate nodes deployed in T_2,T_3, ..., T_M. So, the compromise of node u results clones with all these master keys. Consequently, any clone u*, deployed across the network will be able to authenticate itself to all sensors deployed at T_2, T_3, ..., T_M. Therefore, clones can easily establish PK and infiltrate the network (indirect cloning).

Note that LEAP authors have proposed a solution against the cloning attack in the case of a compromise before T_{est}. In this solution, they rely on the broadcast by the BS of an authenticated list of added nodes, and this using µTesla [6]. In this case, even if the network nodes establish PK with clones, they can revoke them by receiving the list of added nodes.

Although LEAP uses the BS as a trusted third party that broadcasts the list of added nodes, the vulnerability of the cloning attack still exists. The origin of this vulnerability is the lack of mutual authentication between nodes. Indeed, when an attacker compromises a node all constructed clones will have the same identity of the compromised node. In addition, the establishment of PK is based on the identity of nodes; it cannot be done with a clone that has an identity x and keys of node y. Therefore, the broadcast of the list of added nodes identities eventually brings nothing. In other words, as the compromised node is legitimate, so their identity and keys are legitimate also, except that clones are everywhere in the network. The problem is that the network will never identify them.

B. OTMK

Like LEAP, Opaque Transitory Master Key (OTMK) [2] is based on a primary key. However, the PK established between two nodes will not depend on the primary key; it will be randomly generated by both nodes. This ensures the security of existing links in case of the compromise of the primary key. To do this, each node is pre-loaded with a primary key K_{IN}, and the PK establishment is done as follows:

$$u \rightarrow * : \text{Join}|E_{K_{IN}}(u|n_u) \quad (5)$$

$$v \rightarrow u : \text{Reply}|E_{K_{IN}}(v|n_u + 1|K_{uv}) \quad (6)$$

For the maintenance operation, Deng et al. proposed a mechanism, which is also based on the decomposition of the life of the network into a number of time intervals. Each node is pre-loaded with an authentication key H. This key is obtained by applying a chain of one-way hash function $< H_k, H_{k-1}, \ldots H_1, H_0 >$, where each key H_i corresponds to a time interval i. Nodes deployed in the same interval establish their PK as described previously (5 and 6), where the authentication key H_i plays the role of the primary key K_{IN}. Subsequently, each node v calculates its primary key $K_v = \text{MAC}(H_i, v)$ and calculates $H_{i-1} = f(H_i)$ and then removes H_i.

Let u be a new node and v an old node. The node u is pre-loaded with H_j and the node v with H_i (with j> i). When the node v receives the message JOIN from the node u, it responds with a message containing its identity v, a nonce n_v, the index i of the primary key and the associated MAC:

$$u \rightarrow * : \text{Join} \mid n_u \mid u \quad (7)$$

$$v \rightarrow u : v \mid n_v \mid i \mid \text{MAC}(K_v, n_u|n_v|v) \quad (8)$$

Node u can compute H_i from H_j and the index i. Thus, it can generate K_v with H_i and check the MAC, and therefore authenticate v. Since v is authenticated, node u must also authenticate itself to v:

$$u \rightarrow v : u \mid E_{K_v}(K_{uv})| \text{MAC}(H_j, n_v|u|K_{uv}) \quad (9)$$

Once PK is established, u broadcasts its primary key H_j. Therefore, v can authenticate u:

$$u \rightarrow * : H_j \quad (10)$$

Discussion

Although an adversary who compromises K_{IN} cannot compromise already established PK, he can intercept reply messages (6) and then disclose PK. Moreover, this protocol is not secure since listening to the traffic allows an adversary to save the messages exchanged during a time interval j (9), and then, decrypt them after the disclosure of the key H_j (10), which will cause the loss of the confidentiality in the network.

Regarding the cloning attack, an attacker can carry out a direct cloning attack against existing nodes through clones obtained by the compromise of a new node before it erases the primary key H_j. In addition, even if the compromise would be done after clearing H_j, the attacker can (through the index i and the corresponding primary key) conduct an indirect cloning attack against nodes added in the future time intervals T_i (i>j), and this, by replaying the same authentication (8).

C. TBMK

The probabilistic scheme Time-Based Key Management protocol for WSN (TBMK) [3] assumes that the time T_{est}

required to establish keys between nodes is greater than T_{min}, the time required to compromise a node. To reduce the impact of compromising the primary key K_I, authors have decompose the lifetime of the network in P time intervals T_i (corresponding to maintenance phases) when for each time interval corresponds a primary key. In addition, authors used a probabilistic distribution in the pre-loading of sensors, as used in [8].

To do this, the BS pre-loads nodes with a primary key K_i^k corresponding to their deployment time interval k, and a set of m random primary keys corresponding to the future intervals i ($K_{ui} = f_{K_{Ii}}(u)$). After deployment, like LEAP, the first key establishment (corresponding to the time interval T_1) is done through the key K_{I1}. A node u computes its key and broadcasts a message containing its identity and a nonce n_u. A node v responds with a message containing its identity and the MAC of $(n_u|v)$:

$$u \rightarrow * : u \mid n_u \quad (11)$$

$$v \rightarrow u : v \mid MAC(K_{v1}, n_u|v) \quad (12)$$

Having the key K_{I1}, a node u can generate the primary key of v and thus authenticate it. The PK is calculated as follows: $K_{uv} = f_{K_{v1}}(u)$.

For new nodes joining, those deployed at the same time will be able to establish PK as they have the same primary key. They can also establish PK with those deployed in previous time intervals if old nodes have among the randomly pre-loaded primary keys, a primary key derived from the key of the current interval. Fig. 2 presents an example given in [3]. N_n represents a group of nodes deployed at the time interval T_n. The group N_1 is pre-loaded with the primary key K_{I1} and 3 randomly selected primary keys (K_{U2}, K_{U5}, K_{U7}). They can establish PK between them, because they all have K_{I1}. As they can establish PK with the

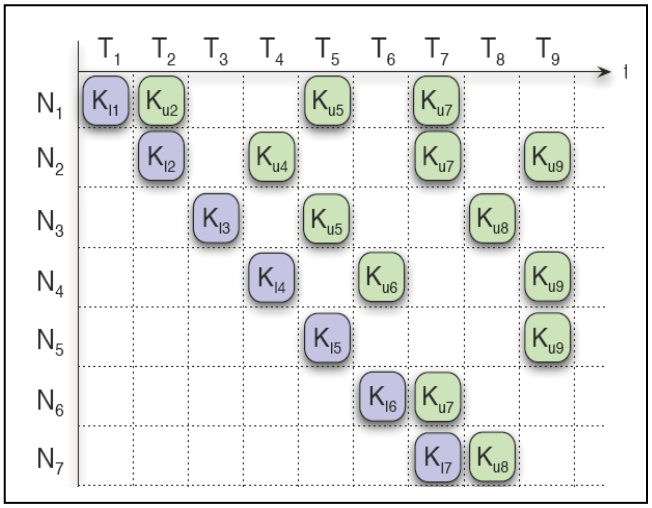


Figure 2. Example of probabilistic primary key pre-loading.

groups N_2, N_5 and N_7 because they have K_{U2}, K_{U5} and K_{U7}, respectively. Groups N_2, N_5 and N_7 can authenticate nodes belonging to the group N1 thanks to their respective primary key K_{I2}, K_{I5} and K_{I7}.

Discussion

Regarding the cloning attack, the compromise of a sensor that belongs to N4 can recover the primary key of the current interval K_{I4} and the primary keys K_{U6} and K_{U9} derived from the primary key K_{I6} and K_{I9}, respectively. Thus, clones can, thanks to the primary key K_{I4}, establish PK with nodes deployed during the same time interval (i.e., nodes of N4) as well as those deployed during the interval T_2 (i.e., nodes of N2), which represent a direct cloning attack.

Even if the compromise was made after the erase of the primary key K_{I4}, primary keys K_{U6} and K_{U9} will never be erased, which will allow clones to establish PK with nodes to be deployed later during the time interval T_6 and T_9 (i.e., nodes of groups N_6 and N_9), which represents an indirect cloning attack.

D. TEKM

Towards Enhanced Key Management (TEKM) scheme [4] is based on the fact that a node u can live up to G_w generations (generation window), which correspond to the addition of new nodes. Before deployment, a node u belonging to the generation j is pre-loaded with a set key KR_j called "KeyRing" containing the primary key IK_j of the generation j and (Gw-1) hidden primary keys $K_{j,l}$ of the future generations ($KR_j = \{IK_j, K_{j,l}\}$). The hidden primary keys are calculated using a secure hash function:

$K_{j,l} = H(IK_l|j)$.

Such as : $j + 1 \leq l \leq j + G_w - 1$.

After deployment, nodes of a generation j have the same primary key IK_j, so it will be used (before being deleted) to establish the PK ($K_{u,v}^j$).

$$u \rightarrow * : u \mid j \mid n_u \quad (13)$$

$$v \rightarrow u : v \mid MAC(K_v^j, u|v) \quad (14)$$

where: $K_v^j = f_{IK_j}(v)$ (15)

$$K_{u,v}^j = \begin{cases} f_{K_u^j}(v) \text{ si } u < v \\ f_{K_v^j}(u) \text{si } u > v \end{cases} \quad (16)$$

For nodes that do not belong to the same generation, PK is calculated differently. Indeed, let u and v be two nodes that belong to the generation g and h respectively ($1 \leq g < h \leq g + Gw - 1$). After deploying the generation h, the PK K_{uv}^{gh} between u (old) and v (new) will be calculated as follows: $K_{uv}^{gh} = f_{K_{gh}}(u|v)$. The node u is already pre-loaded with the key K_{gh}, since its "KeyRing" is as follows: $KR_g = \{IK_g, K_{g,g+1}, \dots, K_{g,h}, \dots, K_{g,g+Gw-1}\}$. For the node v, K_{gh} is the hidden primary key of the group deployed in the

TABLE I. COMPARISON BETWEEN THE STUDIED KEY MANAGEMENT SCHEMES

	Using primary key in		Security			Performance		
	Authentication	PK establishment	Resilience	Cloning		Memory	MAC	Transmission
				Direct	Indirect			
Basic LEAP	Yes	Yes	Good Bad*	Yes* No	Yes	1 key (K_{IN})	Yes	2 local transmissions * d + 1 broadcast from the BS
Extended LEAP	Yes	Yes	Good Bad *	Yes* No	Yes	1 key (K^i_{IN})+ (M-i) keys	Yes	
OTMK	Yes	No	Good Bad *	Yes* No	Yes	1 key (Hi)	Yes	2 local transmissions * d
TBKM	Yes	Yes	Good Bad *	Yes* No	Yes	1 key (K^i_{IN}) + m random keys	Yes	2 local transmissions * d + Proxy
TEKM	Yes	Yes	Good Bad *	Yes* No	Yes	G_w keys	Yes	2 local transmissions * d

*. When the primary key is compromised. d. the average number of neighbors.

previous generation g, and it's not in its "KeyRing". However, it can get it by calculating $K_{gh} = H(IK_h|g)$ as it holds the key IK_h of its generation.

Discussion

When the attacker compromise a new node v belonging to the generation h, and if he recovers the initial key before it is erased, he can use it to calculate the hidden primary key ($K_{xh} = H(IK_h|x)$) of a future generations x, and so to calculate PK of nodes belonging to previous generations (direct cloning).

Moreover, even if the primary key is erased, it can use the hidden primary key contained in its memory to calculate PK of nodes of future generations (indirect cloning).

IV. GENERAL DISCUSSION

The physical compromise of sensors can be exploited in two ways: (1) by disclosing the shared secret used by network nodes to setup PK, thus the existing links can be compromised (e.g., the K_{IN} in basic LEAP is used to calculate all PK) or (2) by reproducing clones that will allow an attacker to infiltrate the network and then to conduct other types of attacks. Both exploits seriously degrade the resilience of a key management scheme. Recall that a perfect resilience refers to the ability of a key management solution to prevent an attacker to not compromise any further communication link other than those involving compromised nodes.

Table 1 compares the discussed protocols in the previous section with respect to security and performance metrics. All schemes use a primary key in the authentication process, so if this key will be compromised, this will allow the attacker to insert malicious nodes. So, all these schemes are vulnerable to the cloning attack, especially the indirect cloning attack. In addition, in all schemes (except OTMK), the PK are also based on the primary key. So, its compromise leads to disclose of all PK already established, which weighs heavily on confidentiality. In LEAP if an attacker succeeds to compromise a node deployed in a time interval before T_{min}, then confidentiality is not guaranteed since the attacker can compute master keys of nodes deployed in the same time interval. In addition, if the attacker compromises a node after T_{min}, it can recover master keys of the future time intervals. Thus, it can clone it and all copies will be able to present themselves as legitimate nodes to new nodes even if they cannot communicate with the existing nodes of the same time interval.

Although the protocol OTMK is very complex, it is not secure, since listening to the traffic allows an attacker to back up the encrypted messages exchanged during a time-slot j and decipher them after disclosure of the key Hj. On the other hand, it can lead an attack cloning by replaying the same authentication through its own nodes in several places in the network.

The problem of time-based key management schemes, regarding the cloning attack, lies on the authentication mechanism between the nodes. This mechanism is generally based on some secret information pre-loaded on sensors to build confidence between nodes. Unfortunately, this latter is set to fail by the physical compromise of nodes, allowing the disclosure of the shared secret. Being aware of this vulnerability and wanting to keep the distributed aspect of their key management schemes, works [1], [2], [3], [4], [14], [9] introduced techniques such as time intervals, hash function, digital signature, etc. to limit the impact of the compromising of sensors. However, these solutions are still vulnerable to the cloning attack as shown in the previous section, even if their assumptions are satisfied.

V. CONCLUSION

In this paper, we presented the weaknesses of time-based protocols, especially regarding the maintenance operation, which is a critical phase in the life of a WSN. The problem lies in the legitimacy of joining new nodes to the network on one hand, and on the other hand, to the physical compromise of sensors that allows an attacker to have clones and inject them into the network. Therefore, the problem is difficult, since new nodes might need to legitimately join the network at times, raising the question of how to perform authentication while defending against cloning attacks. Thus, it is essential to involve the BS or to depend the primary key to the geographic position of nodes to ensure perfect resilience during the maintenance phase.

 ISBN: 978-1-61208-296-7

REFERENCES

[1] S. Zhu, S. Setia, and S. Jajodia, "Leap+: Efficient security mechanisms for large-scale distributed sensor networks," ACM Trans. Sen. Netw., 2002, 2(4), pp. 500-528.

[2] J. Deng, C. Hartung, R. Han, and S. Mishra, "A practical study of transitory master key establishment for wireless sensor networks," Proc. of the 1st International Conference on Security and Privacy for Emerging Areas in Communications Networks (SecureComm 05), Athens, Greece, September 2005. pp. 289–299.

[3] J. Jang, T. Kwon, and J. Song, "A time-based key management protocol for wireless sensor networks," in Information Security Practice and Experience, 2007, pp. 314–328.

[4] B. Tian, S. Han, L. Liu, S. Khadem, and S. Parvin, "Towards enhanced key management in multi-phase ZigBee network architecture," in computer communication, 2012. 35(5): pp. 579-588.

[5] A. Fanian, M. Berenjkoub, H. saidi, and T. A. Gulliver, "A high performance and intrinsically secure key establishment protocol for wireless sensor networks," Comput. Netw, 2011, vol. 55, pp. 1849-1863, doi:10.1012/j.comnet.2011.01.012.

[6] A. Perrig, R. Szewczyk, V, Wen, D, Cullar, and J. D. Tygar, "SPINS: Security protocols for sensor networks," Proc. of the 7th annual ACM/IEEE international conference on mobile computing and networking, July 2001, pp. 189–199.

[7] Y. S. Han, W. Du, S. Chen, and P. K. Varshney, "A key management scheme for wireless sensor networks using deployment knowledge," Proc of IEEE INFOCOM04. Hong Kong: IEEE Press; 2004. pp. 586–597.

[8] L. Eschenauer and B. Gligor, "A key-management scheme for distributed sensor networks," Proc of the 9th ACM conference on computer and communication security, Washington, DC, USA, 2002. pp. 41–47.

[9] S. Blackshear, M. Rakesh, and Verma, "R-LEAP+: Randomizing LEAP+ key distribution to resist replay and jamming attacks," in: SAC'10 March 22, 2010, Sierre, Switzerland.

[10] D. Manivannan and pp. Neelamegam, "WSN: Key issues in key management schemes-A review," Research Journal of Applied Sciences, Engineering and Technology 4(18): 3188-3200, 2012.

[11] W. Du, J. Deng, Y. S. Han, P. K. Varshney, J. Katz, and A. Khalili, "A pairwise key pre-distribution scheme for wireless sensor networks," ACM T. Inform. Syst. Se., 2005. 8(2): pp. 228-258.

[12] J, Zhang, and V. Varadharajan, "WSN key management and taxonomy," Journal of Network and Computer Application 33 (2010) pp. 23-75.

[13] B. Parno, A. Perrig, and V. Gligor, "Distributed detection of node replication attacks in sensor networks," Proc. of the 22th IEEE symposium on security and privacy (S&P'05); May. 2005. p. 49–23.

[14] O. Delgado-Mohatar, A. Fúster-Sabater, and J. M. Sierra, "A light-weight authentication scheme for wireless sensor networks," Elsivier, Ad Hoc Networks 9 (2011) pp. 727–735.

 ISBN: 978-1-61208-296-7

A Formal Method for the Evaluation of Component-based Embedded Systems: Application to Technical Choices for CSTBox Toolkit

Daniel Cheung-Foo-Wo, Éric Pascual
Centre Scientifique et Technique du Bâtiment (CSTB)
Sophia Antipolis, France
daniel.cheung@cstb.fr, eric.pascual@cstb.fr

***Abstract*—When designing applications for sensor or actuator network management, deployed networks of equipments often use an underlying component-oriented approach, such as OSGi (formerly Open Services Gateway Initiative, now OSGi Alliance), which provides flexibility in terms of software development and maintenance. It also leads to combining sensors and actuators from different manufacturers using different protocols, libraries, and communication media. To make the interface between components compatible, additional software layers are implemented on top of heterogeneous libraries and protocols. Our intention in this paper is to confront our OSGi applied software solution coping with these heterogeneity and flexibility to contextual timing constraints in order to propose solutions in terms of OSGi extensions and generalization. We formally describe the performance function of the connectors of our toolkit called CSTBox (CSTB Sensing Telemonitoring Box) faced to soft real-time constraints derived from real-world deployments of European projects in domains such as ambient assisted living for the elderly and energy use in offices according to a simple classification of underlying middleware in terms of performance. Running applications, result analyzes, and timing measurements are performed on the SOC (System on Chip) card Raspberry Pi hardware for low-consumption and other embedded systems like routers and gateways.**

Keywords-component; soft real-time constraints; OSGi; D-Bus; sensor network data processing measurement;

I. INTRODUCTION

The problem addressed in this paper is derived from research projects [1]–[3] in ambient instrumentation of buildings. The objective is to provide software managing networks of equipments in the form of an OSGi-based toolkit for embedded systems in ambient assisted living and analysis of energy use. However, these domains impose some requirements on systems. Firstly, in order to protect *privacy* in home activity monitoring, external data transfer is securely kept minimal. Besides, the system must be robust to network failures to avoid information loss. As a result, most of the data processing is done in-place, which promotes the systems from simple data acquisition and transfer units to full processing units. Secondly, these domains concern monitoring for gathering statistical information and require sufficient *soft real-time* (RT) constraints on observation rate and reaction delay. Various applications from assisted living and energy use accordingly *share* a great part of functionalities, and emphasize the need for a component-based solution. The system must also be flexible in supporting multiple sensor networks connecting from local and remote sites. Furthermore, the hardware equipments, such as sensors, actuators and network coordinators, might differ because of their origin and their protocol. These *heterogeneous* equipments are thereby mixed for a given instrumentation of buildings (see Figure 1).

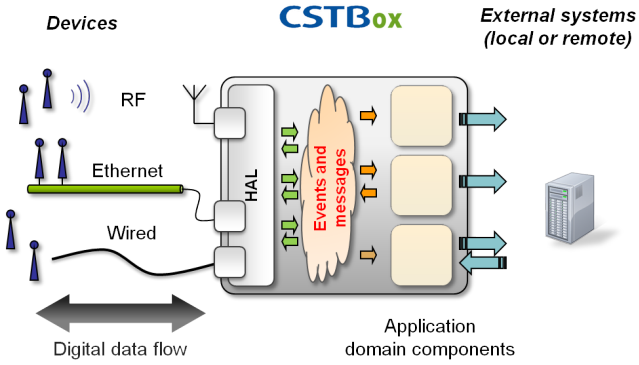


Figure 1. Operation principle of the CSTBox abstraction mechanism

On account of these requirements, our software toolkit is characterized by its assembly of soft RT components using heterogeneous equipments. We proposed a solution [4] tackling sensor-actuator heterogeneity by means of representing an *equipment* (decomposed into parts) as a set of readable or writable *state variables* associated with a *type*, an *identifier* of the associated part and the *state* of the equipment part.

This paper deals with the *confrontation of this OSGi approach to timing constraints induced by assisted-living and energy-use domains* and its consequences on the choice and the design of the component communication provided by the underlying OSGi platform, i.e., *software connectors*. The remainder of this paper is organized as follows: next section presents some related works. Section III introduces the model of our toolkit and presents a formal approach for scoring its performance. In Section IV, we detail the problem of applying constraints and the way it addresses the issue of the proper selection of connectors. We conclude in Section V on perspectives.

 ISBN: 978-1-61208-296-7

II. Related work

Software connector design and choice can be studied in several ways. For instance, connector choice can be achieved by querying an existing classification of connectors [5] organized in terms of the following six main dimensions: the amount of transferred data, delivery intervals, performance requirements, the number of consumers, access policies, and geographic distribution of components. A data distribution scenario is then expressed as a query against the above dimensions. The selection algorithm is thus tasked with deciding what available connectors will suit a given scenario, which is accomplished by maintaining a knowledge base of connector profiles. A connector profile contains metadata including locality, number of providers/subscribers, throughput, delivery semantics, and routing. The authors' focus is on automatizing the selection algorithm for fetching software connectors based on the key dimensions. Although automation represents a great benefit to software engineering, we are more interested in a formal tool helping the developer figuring out the adequate use of connectors.

Another related area of research relies on software architecture programming dealing with frequent updates. Reference [6] presents the design of connectors by combining *white-box* components which semantics and routing are known. The authors focus on mathematical properties, such as associativity of the combination, forming an algebraic way of designing connectors. They use this algebra to integrate, swap or remove functionalities expressed in terms of components. They consider that the set of all adaptations that may be deployed cannot be anticipated at design time, which is beyond our requirements. In such case, adaptations may interfere when they are combined. However, they model the assembly of components as a directed graph (data-flow) providing them with a formal tool to identify properties and errors. We consider designing a formal model compatible with this graph representation and algebra.

Finally, in a software engineering point of view, OSGi provides all the requirements expressed in the Section I. However, based on a qualitative experience, [7][8] decided not to use OSGi as a connector implementation for three reasons. Firstly, existing good quality legacy components shall not be rewritten in Java for costs saving. Secondly, components of the system should sometimes be written by specialists who have their specific language of domain (e.g., Matlab). Thirdly, unless distributed, OSGi runs the application in a single virtual machine, i.e., every bundle shares the same memory area, so the lowest quality bundle determines the quality of the entire system. Consequently, they choose D-Bus (Desktop Bus) which is an inter-process communication solution with the following claimed properties: language independence and efficiency (however, no quantitative results is given for that case). Hence, components become separate processes directly managed by the operating system benefiting from its RT properties if available. D-Bus is heavily used in GNU/Linux both in desktop and system applications, and it is also ported to other operating systems including Windows. There exists bindings for a wide range of programming languages. Furthermore, it is also possible to connect several D-Buses running on different computers making distributed computing possible, which becomes in competition with initiatives like DOSGi [9] (Distributed OSGi). As a result, a D-Bus-based toolkit provides two new characteristics: language independence and component separation (see Section IV-B).

As a result, although these approaches propose complete solutions for component connection abstract and concrete model, none considers applying it to the widely used OSGi concrete model and perform a quantitative study of its incremental extension and provide a systematic methodology to score existing solutions to ensure the most adequate solutions rise to the top, which is the focus of this paper.

III. Constrained model of components

To present our connector model, we introduce in this section the model of our component system. We abstract it from any specific notation and represent it as a directed graph as in [6].

Definition 1. (Component graph) *A component graph $\mathcal{G}$ is a set of vertices V and directed edges E, noted $\mathcal{G} = (V, E)$. A vertex v has the attribute $v.\textbf{\textit{typ}}$, a type. An edge e from v_1 to v_2 of V is written as $e = (v_1, v_2, l)$ where l is a label.*

Each vertex has a type which depends on the underlying system, i.e., communication bus or programming language.

Definition 2. (Input and Output) *Let $\mathcal{G} = (V, E)$ be a directed graph. For each vertex $v \in V$, its inputs are labels $I_1, \ldots, I_m$ noted $v.\boldsymbol{I}$ such that $\forall k \in [1, m], (v_k, v, I_k) \in E$. Similarly, its outputs are $O_1, \ldots, O_n$, noted $v.\boldsymbol{O}$ such that $\forall k \in [1, n], (v, v'_k, O_k) \in E$.*

In this model, a software connector is simply represented by a vertex as a component.

A. Updating state variables

The state variables, as introduced in Section I for solving heterogeneity problems, are represented by a set $\mathcal{S}$ where each element $\sigma \in \mathcal{S}$ has the attributes $\sigma.\textbf{sta}$, its state, $\sigma.\textbf{id}$, its identifier, and $\sigma.\textbf{typ}$, its type.

To update state variables, we define a fixed and contextual common message structure. The fixed part contains the *time stamp* of the message, the *type* of the associated state variable, a *regular expression (regexp) id* matching one or several state variable ids, and the *state*. The contextual part is a key-value *map* function on contextual information, such as the unit of the state, etc. Special connectors implementing the hardware abstraction layer are exhibited from the component graph. We call them HAC connectors (Definition 3).

 ISBN: 978-1-61208-296-7

Definition 3. (HAC) *Let $\mathcal{G} = (V, E)$ be a directed graph. A hardware abstraction connector (HAC) is a vertex $\gamma \in V$ such that for all $\pi \in \{\boldsymbol{I}, \boldsymbol{O}\}$, $\gamma.\pi = \gamma$ has the attributes **ts** (time stamp), **sta** (state), **re** (regexp id), **typ** (type), and **m** (key-value map).*

HAC connectors translate a regexp id from its input into a constant string id. On the one hand, as collected data are related to physical events, i.e., measurements of quantities (temperature, energy consumption, etc.) or notifications of state change (door opening, motion detection, etc.), the input regexp id is then reduced to a constant string representing the id of the state variable (e.g., *living-room.light.ceiling*). On the other hand, a sub-assembly of components should compute the matching of the state variable handling the action and translate the command to the proper communication frames. It allows us to specify the target using pattern matching rules, such as *living-room.light.** (id is supposed to follow a predefined format), which will alter state variables matching the rule receiving the appropriate command.

Property 1. (HAC invariant) *For all HAC noted γ in a component graph, the state of γ reflects the state of one or several state variables according to the matching of γ's regexp, i.e., for all HAC $\gamma \in V$ and $\sigma \in \mathcal{S}$, if $\sigma.\textbf{typ} = \gamma.\textbf{typ}$ and $\sigma.\textbf{id} \cong \gamma.\textbf{re}$, then $\sigma.\textbf{sta} = \gamma.\textbf{sta}$ where the symbol '$\cong$' stands for a regexp matching function.*

Definition 1 highlights the flexibility benefited from HAC connectors to target multiple state variables. Such an approach of the communication between the building blocks of the application allows us to decouple component implementations. We presented above the model of the toolkit and its properties. We propose in the next paragraphs one way of applying constraints to express a notion of performance.

B. Soft real-time constraints

Soft RT constraints impose that system response time should be adequate in average. At the connector level, this means that its delays should be compatible with these constraints. Thus, we propose to work with the ratio τ (called *performance*) of the duration noted d_C of a function call in C language as our yardstick reference to be independent of the CPU (central processing unit) load, i.e., the time between the function is invoked in a source component and the execution of its first instruction in a target component, over its own duration. We choose C for its widespread use as a simple procedural abstraction. This ratio is computed experimentally (see Section IV).

Definition 4. (Connector performance) *The connector performance τ of a connector type $v.\textbf{typ}$ is a function $\tau(v.\textbf{typ}) = \frac{d_C}{d}$ scoring its performance between 0 (low) and 1 (high performance) where d is the computation time of $v.\textbf{typ}$.*

To compute the overall performance of the connectors of the system, we propose to take the average sum of all connector performances on the path of the predominant control flows. We do not seek to provide a predictive calculus for a given application, but rather a metric comparison between the different systems, achieved by a simple arithmetic average without weighted control flow.

Definition 5. (Performance of system) *The performance P of a set of n types of connector $(t_1, \ldots, t_n)$ and control-flow paths $\mathcal{F}$ is:*

$$P = \frac{1}{|\mathcal{F}|} \sum_{f \in \mathcal{F}} \left(\frac{1}{\sum_{j \in [1,n]} k_{f,j}} \sum_{i \in [1,n]} k_{f,i}.\tau(t_i) \right) \quad (1)$$

where $k_{f,i}$ is the number of times the path f traverses a connector of type t_i and $|\mathcal{F}|$, the cardinal of $\mathcal{F}$.

We built a model with a global performance function (Equation 1) which purpose is to provide a way to score the performance of a middleware. We show in the next section how to apply it to evaluate development solutions.

IV. D-Bus as an OSGi middleware extension

A. Connectors in iPOJO/OSGi

Apache Felix was selected as the OSGi implementation and iPOJO [10] (injected Plain Old Java Object) is used on top of it to provide an extensible and declarative component-oriented model. Thus, additional "handlers", i.e., *connectors*, can be implemented by means of extensions of the *PrimitiveHandler* class to intercept access to attributes and method invocations of *pojos*, i.e., functional codes. The first requirement in our projects was to specify service filters in a configuration file that selected the components associated with a required service. The existing connector *Require handler* of iPOJO proposes only a design-time service filter. We created a new handler (called *VariableRequire*) based on the latter to read a configurable property initialized with a given configuration file before adding filtering information during the assignment of a component. Secondly, we created another handler called *RegexpSubscriber* extending the event *Subscriber* of iPOJO to listen to broadcasted OSGi events. It provided a regexp HAC connector matching the OSGi topic as a complementary information instead of the internal OSGi specific filter. We do not detail the implementation of both connectors in this paper, but we focus on scoring them to provide a quantitative analysis of the system performance. We measured the C method call duration $d_C = 7\mu$s on average on Raspberry Pi (240 measures during 4h on low CPU load $< 1\%$) by using *gettime()* function. Two control flows are implemented. The first one writes sensor values into the database and traverses two connector types: 1 HAC (type t_0) and 2 iPOJO calls (type t_1). The second one logs battery level traversing the same connector types. Performance computation is shown in Table I.

 ISBN: 978-1-61208-296-7

B. Towards D-Bus extension

We introduced D-Bus in our architecture as in [7], in particular, by incrementally replacing the OSGi middleware. The idea was to gradually put the language-agnostic D-Bus (or other effective middleware) at the center of the communication between software components (see Figure 2) and observe the evolution in performance.

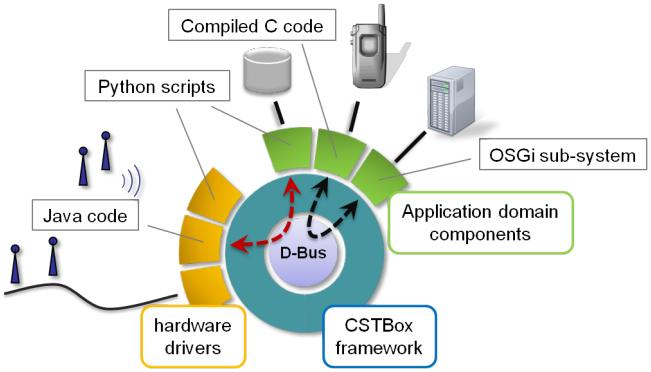


Figure 2. The operation principle of the CSTBox abstraction mechanism

D-Bus implementation requires a single process per bus entity, manages the process interface registration, and keeps overlaps in naming coherent by queuing bus requests. It also manages the life-cycle of processes. We have implemented the same sub-assembly as in Section IV-A, but in Python and D-Bus. The control flows are the same and traverse 2 types of connectors: 1 HAC (t'_0) and 2 Python calls (t'_1).

Table I
OSGi/iPOJO AND D-BUS PERFORMANCE COMPARISON

1 HAC (RegexpSubscriber OSGi)	$\tau(t_0) = \frac{7\mu S}{46000\mu S} \simeq 0.02\%$
2 iPOJO service call connectors	$\tau(t_1) = \frac{7\mu S}{13749\mu S} \simeq 0.05\%$

Performance of OSGi: $P_{\text{OSGi}} = \frac{1}{3}\left(\tau(t_0) + 2\tau(t_1)\right) \simeq$ **0.04%**

1 HAC (CSTBoxEvent D-Bus)	$\tau(t'_0) = \frac{7\mu S}{4061\mu S} \simeq 0.17\%$
2 Python method call connectors	$\tau(t'_1) = \frac{7\mu S}{31\mu S} \simeq 22.58\%$

Performance of D-Bus: $P_{\text{D-Bus}} = \frac{1}{3}\left(\tau(t'_0) + 2\tau(t'_1)\right) \simeq$ **15%**

Table I qualitatively shows that the performance of D-Bus is greater than OSGi's for this simple type of application architecture. This approach provides us with a tool to evaluate engineering choices and converge to suitable solutions, even though it still needs more experiments and validations.

V. CONCLUSION

The work described in this paper is derived from our experience designing, installing, and running experimentation for sensor network management and data processing unit engineering as an OSGi component toolkit. The design/verification flow is currently implemented as a helper tool computing performance from architecture description independently of the networks of equipments. Our toolkit provides and maintains a rich interfacing system, i.e., set of connectors between components. We worked at extracting the practical advantages from experiments and we capitalized on it by generalizing the connector model and evaluation method. As perspectives, we are considering achieving a complete experimental model of performance and the design of RT connectors and their certification by means of program proof in suited languages [11] to ensure safety property.

REFERENCES

[1] N. Zouba et al., "A computer system to monitor older adults at home: Preliminary results," Gerontechnology Journal, vol. 8, no. 3, July 2009, pp. 129–139.

[2] SIGAAL Consortium, "SIGAAL Project," http://sigaal.org, [retrieved: June, 2013].

[3] A. Flamand and N. Roudil, "Domestic energy consumption: Inhabitants practices, rationality and motivation," in Int. Scientific Conf. on Sustainable Consumption - Towards Action and Impact, Abstract, Hamburg, Germany, Nov. 2011, p. 146.

[4] A. Zarli, E. Pascual, and D. Cheung, "Information and communication technology for intelligent and integrated control in buildings: Current developments and future research," in 27th Int. Conf. on Applications in IT in the AEC Industry, CIB, 2010 [published online, 10 pages].

[5] C. A. Mattmann, D. Woollard, and N. Medvidovic, "Exploiting connector knowledge to efficiently disseminate highly voluminous data sets," in 3rd Int. Workshop on Sharing and reusing architectural knowledge. New York, NY, USA: ACM, 2008, pp. 37–40.

[6] S. Fathallah, S. Lavirotte, J.-Y. Tigli, G. Rey, and M. Riveill, "The dynamic composition of independent adaptations including interferences management," in 7th Int. Conf. on Soft. Eng. Advances, Lisbon, Portugal, Nov. 2012, pp. 678–684.

[7] P. Hanák et al., "System architecture for home health and patient activity monitoring," in 5th European IFMBE Conf., 2011, pp. 945–948.

[8] T. Kovácsházy, G. Fodor, and C. B. Seres, "A distributed power consumption measurement system and its applications," in 12th Int. Carpathian Control Conf., May 2011, pp. 224–229.

[9] S. Mohammed, D. Servos, and J. Fiaidhi, "Developing a secure distributed OSGi cloud computing infrastructure for sharing health records," in 2nd Int. Conf. on Autonomous and Intelligent Systems. Berlin, Heidelberg: Springer-Verlag, 2011, pp. 241–252.

[10] C. Escoffier and R. S. Hall, "Dynamically adaptable applications with iPOJO service components," in 6th Int. Symp. on Software Composition. Braga, Portugal: Springer-Verlag, July 2007, pp. 113–128.

[11] P. Letouzey, "Extraction in Coq, an overview," in Logic and Theory of Algorithms, 4th Conf. on Computability in Europe, A. Beckmann, C. Dimitracopoulos, and B. Löwe, Eds., vol. 5028. Springer-Verlag, 2008, pp. 359–369.

 ISBN: 978-1-61208-296-7

Environmental Monitoring based on Wireless Sensor Network via Mobile Phone

Laura M. Rodríguez Peralta
Cómputo Académico, Universidad Popular Autónoma de Puebla (UPAEP), Puebla, México.
E-mail: lauramargarita.rodriguez01@upaep.mx

Andrea M. M. Abreu, Lina M. P. L. Brito,
Competence Centre of Exact Sciences and Engineering, University of Madeira (UMa), Madeira, Portugal.
E-mail: nickandrea8@gmail.com, lina@uma.pt

Abstract - **Preventive conservation of artwork is a vital issue for museums. Therefore, it is fundamental to continuously monitor its environment. To accomplish this, the WISE-MUSE project proposes the use of Wireless Sensor Networks (WSNs) for monitoring and automatically control museums' environment and structural health. The deployment of a Wireless Sensor Network can help to implement these measurements continuously, in a real-time basis, and in a much easier and cheaper way than when using traditional measuring equipments and procedures. Nevertheless, continuous monitoring implies that the persons in charge for the museum must be immediately informed of any abnormal value detected in its environment. Thus, this contribution arises: the development of a new mobile application that allows users to continuously monitor the environment and be aware of any undesirable changes that may occur**.

Keywords-wireless sensor networks; mobile applications; environmental monitoring; preventive artwork conservation.

I. INTRODUCTION

Controlling some of the major causes of deterioration and damage of artwork in a museum environment avoids or reduces the need for invasive conservation treatment and ensures that artwork is protected. Therefore, preventive conservation of art in museums is of extreme importance.

Preventive conservation is achieved mainly by controlling the indoor environment, especially temperature, humidity, and light levels. However, environmental monitoring and control in this type of buildings must respect some specific requirements due to esthetical reasons, like minimizing the visual impact caused by the installation of monitoring systems. Besides, the current economic situation requires the choice of a cheap solution.

The deployment of a Wireless Sensor Network (WSN) in a museum can help to implement these measurements automatically and continuously. This solution is much cheaper and easier to install than traditional measuring equipment. Besides, the major advantage of using WSNs in this type of environment is causing very little visual impact due to the small size of sensor nodes, and an extremely important aspect in the context of a museum: being wireless.

Considering these requirements, the WISE-MUSE project [1, 2] was created with the purpose of employing WSNs for monitoring the environment and structural health of museums, and automatically control relevant environmental parameters. The importance of this project, obviously besides monitoring the environment of the museum, is to study and to demonstrate some practical aspects of WSNs, since most of the applications have not been tested in real WSNs.

The WISE-MUSE project is an on-going project that was initially developed for a contemporary art museum that is located on the premises of Fortaleza São Tiago, a fortress located in Madeira Island. Currently, this project has been extended to Madeira Whale Museum, also located in the island.

In this paper, an emphasis will be given to the WISE-MUSE mobile application, which is an important contribution of the WISE-MUSE project that was not planned from the beginning. This contribution came up to meet the needs that have arisen throughout the course of this project.

The WISE-MUSE mobile application was developed to meet the concerns of those responsible for the museum to be always informed about abnormal environmental parameters, when they are not on the premises of the museum. The WISE-MUSE mobile application is developed for Android and iPhone mobile platforms and allows an easy and flexible way to remotely monitor the museum environment.

This paper is organized as follows. Section 2 briefly describes some of the most important aspects of the WISE-MUSE project. In Section 3, an emphasis will be given to the WISE-MUSE mobile monitoring application designed to monitor a WSN through either Android or iPhone mobile platforms. Section 4 describes related work and provides a comparison with the WISE-MUSE mobile application. Section 5 summarizes the most important conclusions, whereas Section 6 points out some directions for future work.

II. THE WISE-MUSE PROJECT: APPLYING WSNS TO MUSEUMS' ENVIRONMENTAL MONITORING

The main goal of the WISE-MUSE project is to create a WSN for monitoring and automatically controlling the most critical environmental parameters of interest to a museum,

 ISBN: 978-1-61208-296-7

which are humidity, temperature, and light, in museums (exposition rooms and storage rooms).

Nevertheless, pollutants are also of major concern for museum managers due to their influence in the degradation of artworks; therefore, this project also involves the monitoring pollutants, namely CO and CO_2. Thus, the major goal of the WISE-MUSE project can be divided into several sub-goals, which comprise:

- To create a WSN for monitoring the most critical environmental parameters in museums (both in exposition and storage rooms).
- To monitor pollutants as they are also a major concern for museum managers.
- To give users and museum' managers the possibility of consulting the collected data in real-time, via different data formats (e.g., graphs, tables, colour gradients, etc.), and consulting the historical reports whenever needed.
- To analyze data in order to verify its compliance with the art conservation rules. If a significant variation (below or above the required levels) on at least one of the measured parameters occurs, an alert has to be automatically sent to the user through a mobile phone (SMS) or e-mail, in order to increase the efficiency of this environmental monitoring system.
- To control the environmental conditions through the automatic control of the air conditioning and dehumidifying systems of the museum.
- To visualize data integrated in a 3D representation of the building.
- To remotely monitor the state of the emergency doors of the museums, giving their managers and security personnel more awareness of their state, and avoiding their displacement to examine the state of each door.
- To implement low-cost devices for monitoring and controlling environmental parameters, as well as the state of emergency doors.

So, this project proposes an environmental monitoring solution that involves installing only one network, using nodes that have the ability to monitor several parameters at the same time. In the context of the WISE-MUSE project, new wireless sensor nodes [1, 2] were developed. These new nodes are smaller and cheaper than the nodes used in other solutions, like the ones used by Spinwave [3]. Besides, the concepts encompassed in WISE-MUSE project are appropriate to be applied either in historic or new buildings. These factors are the basis for making the solution proposed by the WISE-MUSE project more suitable to the environmental monitoring of buildings.

The WISE-MUSE project offers several advantages in relation to other approaches, namely: i) visualizing sensors' data using either a web-based platform or a mobile phone; ii) changing sensors' parameters through the platform; iii) creating and manipulating collaborative sessions using the WISE-MANager tool [4]; and, finally, it allows to monitor the state of the WSN through the awareness tool or the WISE-MANager tool [4]. Another advantage is that, using the same visualization platform; it is also possible to monitor the state of the emergency doors and the environment of cold chambers of the museum [5].

The WISE-MUSE project was validated in two real scenarios: i) Museum of Contemporary Art of Funchal and; ii) Madeira Whale Museum. These scenarios were chosen to reinforce the validation since they have completely different characteristics. The Contemporary Art Museum of Funchal, situated in the city of Funchal, on the island of Madeira, Portugal. Fortaleza São Tiago is an irregular-shaped fortress from the 17th century, which today houses the Museum of Contemporary Art of Funchal). On the other hand, Madeira Whale Museum is a recently built museum located in a small fishing village, on Madeira Island. The museum focuses on the whale hunting activities that were part of many fishermen's lives for several decades and houses several types artifacts, like 3D models of whales, objects made of whale bones, etc.

III. THE WISE-MUSE MOBILE APPLICATION.

To complement the monitoring platform already developed and installed in the museum, a mobile application for the environmental monitoring of museums was developed. With this application, the managers of the museum can see data anytime and wherever they are, they can receive notifications of unwanted situations even when they are not in the museum.

This application was implemented in Android, and later it will be extended to the iPhone platform. The reason for using these platforms is because these Operating Systems (OS) are the most widely used nowadays. This application will allow the visualization of data produced by a WSN installed in the Whale Museum. This application will also display in a user-friendly interface data captured by the sensor nodes installed in the exhibition rooms, in cold chambers and the emergency doors, through graphs and maps. The historic data collected can also be found in tables.

Thus, the monitoring system will allow to:

- Observe the state and the most important environmental monitoring information, monitoring of emergency doors and cold chambers;
- Consult the historic of sensed data;
- Visualize data from environmental monitoring, monitoring of emergency doors and cold chambers, through graphs and maps, thereby making it more noticeable;
- Visualize environmental parameters (humidity, light, temperature, CO2 and CO) collected both in exhibition rooms and cold chambers;
- Generate notifications in order to alert the user of a parameter whose value is not within the desired levels.

 ISBN: 978-1-61208-296-7

This application will display the measurements received by wireless sensors, in a fast and reliable way, so that as soon as a problem is detected, those in charge for the museum are informed about it, being able to act in time and solve the problem, preventing major damage.

Use cases detail the functional requirements of the application, thus giving a clear idea of the different activities that users can perform, when using the application.

The most important use cases used for this application can be seen in Figure 1, in which the behavior of two important stakeholders can be seen; these are the client (person responsible for constant monitoring of data from the museum), and WSN coordinator or gateway (system responsible for sending the data required for monitoring).

The WSN coordinator will send data from the various sensors, in real-time manner, to be visualized later by the customer using this application.

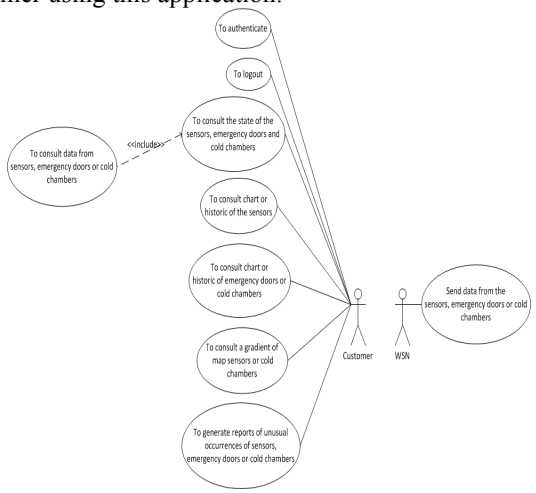


Figure 1. Use case diagram of the WISE-MUSE mobile application

Figure 2 shows the architecture of the WISE-MUSE mobile application. It follows the client-server paradigm.

It is possible to observe that there is a search service that sends the data to the database or gets data form the database, which is implemented through the PHP server, The PHP server, in turn, takes care of the communication between the mobile platform and MySQL database.

Figure 2. Client-server architecture of the WISE-MUSE mobile application

This communication is initiated by the client, which makes a request to the PHP server using the HTTP protocol. This protocol is integrated into the Android operating system. Subsequently, the server connects to the MySQL [6] database using SQL language [7]. The server encrypts data using the JSON format, which will be used by the WISE-MUSE application to obtain and process this data, showing it later on their Android device [8].

Regarding the implementation different mockups were created, in order to show the interaction between the different scenarios, as well as their presentation. A planning scenario was conducted both for the iPhone and Android platforms. In this paper, only the most important navigation scenarios of these two OS are presented.

A. Android platform

Figure 3 shows several scenarios of the Android platform. The application menu (Figure 3a) can be expanded to view some of its options; the menu can be saved after choosing what to visualize, thus allowing optimizing the size of the screen for the presentation of information. Then, it is possible to see the information (Figure 3b) and state (Figure 3c) about the most important sensors, emergency doors and cold chambers, allowing to choose the floor where the different sensors to be visualized are located. The application also allows selecting the different rooms of a previously selected floor, finishing with the visualization afterwards.

Next, the user can select to visualize data by a graph (Figure 3e) or a historic (Figure 3f), selecting different visualization criteria (Figure 3d). First, a date can be chosen; then the floor where sensors are located can be chosen (all, several or a specific floor); subsequently, all rooms may be chosen, several rooms, or only one room, from the floor which has been previously selected; similarly, the device category may be chosen, i.e. the type of device (sensors, emergency doors or cold chambers); besides, it is possible to choose whether to view all or only a few of these sensors. After choosing the sensor, it is then possible to choose the types of parameters to be displayed, the user may opt for one, all or some, among which are: temperature, humidity, light, CO2 and CO. It is also possible to display a graphic or historic with all the selected criteria.

B. iPhone platform

Regarding the iPhone platform, the application menu is always visible and the user can choose one of the options, at any time. When selecting the scenario, it remains green, allowing giving feedback to the user. In Figure 4, the user can visualize several scenarios of the iPhone platform on which data is represented in different maps (Figure 4a). First, the user has to select the type of map (Figure 4b) that user wants to observe, such as the map of temperature, humidity, light, CO2 and CO.

 ISBN: 978-1-61208-296-7

Figure 3. Android Prototypes. a) Application menu. b) Viewing the state of a sensor. c) Display of data from a selected sensor. d) Graphics menu. e) Graphics display. f) Historic display.

Inside the maps, we can choose the criteria according to which data will be visualized, such as the floor, the room and the sensor, and for these last three, it is possible to choose one, all or only some. In the case of temperature map, the visualization criteria are the same; with the exception of have the possibility of choosing the type of device to be visualized (sensor or cold chambers). In any case, the presentation of data is performed using the map (Figure 4b). The user can also see the notifications (Figure 4c) of devices that are in danger or emergency, through its representation using different colors depending on the corresponding status.

When the user performs authentication in the application, if notifications exist, they will be shown to the user (Figure 4c), in order to alert him immediately. Nevertheless, the user may also view the notifications at any time.

IV. RELATED WORK.

When looking for other environmental monitoring works, we realize that there are some applications in either software or hardware that can monitor effectively the environmental conditions of a specific place.

The Overseer application (Network Monitoring Software) [12] monitors the environmental status of the room where a server is located, i.e., the user can monitor temperature and humidity of a data center via an EM1 monitor [13]. Alerts are sent to the user via e-mail to the mobile phone if the upper or lower limits of these parameters have been reached.

On the other hand, there are different applications which help in environmental monitoring for instance, EnviroSense [14], which is software that monitors temperature and humidity, being useful for either for agriculture, computer rooms or any area that needs monitoring. This application alerts the user by sending SMS and reports, to a mobile phone.

Among the works found that most resembles the WISE-MUSE application, we have found a surveillance monitoring system developed for the Android mobile platform [12, 14]. This system aims to address the deficiencies of security systems through WSN via the Android mobile phone, allowing the user to monitor a particular place at any time. In this work, the activity sensor was used to display data on humidity, temperature, carbon dioxide and gas concentrations, allowing shutting down or activating the air conditioning equipment, if the sensors detect abnormal data. Additionally, it indicates any unusual movement of material or people outside or inside the room. All this can be observed at the interface of the corresponding application, where data from the different sensors can also be visualized, as well as a video of the room captured by a digital camera via USB.

In the specific case of the project WISE-MUSE, an application was developed with the aim of allowing to monitor the different environmental conditions (temperature, humidity, light, CO_2 and CO) captured in a museum, thus enabling the persons responsible for the conservation of artworks in the museum to detect if the environmental parameters are outside acceptable limits, alerting the museum staff in a fast and efficient manner.

 ISBN: 978-1-61208-296-7

Figure 4. iPhone Prototypes. a) Choice of different maps. b) Map view of humidity gradient. c) Display of notifications.

Other application that has some similarities to our work, is a system that allows monitoring different WSN sensor nodes (temperature and humidity) allowing the users to can visualize these parameters, and send notifications when a certain value has exceeded its limit. These sensors can be visualized by building plan, floor or sensor type; it is also possible to customize the reception of notifications from various sensors that are associated with a particular person [13].

In order not to only send notifications via SMS or email, but also to allow visualizing the data collected by the WSN at any time, the WISE-MUSE application was implemented for a mobile platform. This will ensure efficient monitoring of real-time data in order to allow better preservation of artwork.

Table 1 shows a comparison between the WISE-MUSE mobile application, the application developed by [12, 14] and a mobile application to monitor a WSN [13].

The major advantage of the WISE-MUSE mobile application is to provide complete overview of all real-time data collected by different types of sensors, emergency doors and cold chambers. It also allows visualizing data using graphs, maps and historic.

Both in the security system developed for the Android platform [12, 14] as well as in the mobile application to monitor a WSN [13], sensor data can be visualized only in real-time. That is, they do not let the users to view and analyze historic data through graphs and maps of gradients.

Another advantage of the WISE-MUSE is being possible to observe the different data from one or more sensors, emergency doors and cold chambers at a time. This is not possible with other applications. The application developed by Moreira et *al.* [13] allows selecting from various sensors, not allowing the visualization of the state of emergency doors, whereas with the security system presented in [12, 14] can only be observed data from a sensor.

Regarding the work that focus on surveillance monitoring [12, 14] it was observed that the environmental parameters are displayed on the Android application, but there is nothing to help understanding if the data are abnormal or not, which may difficult the interpretation of data. Due to this, the WISE-MUSE mobile application uses different colors to give feedback to the user and alert them if any parameter is outside the normal values. For this reason, the WISE-MUSE mobile application was designed considering different types of sensors, deployed by several floors and rooms that will need to be monitored.

The uses of cameras placed in locations to record and transmit images via a USB interface is an important advantage of the application proposed in [12, 14]. Thus, the architecture of WISE-MUSE project includes the installation of a camera-based monitoring module. Once this module is installed and integrated in the WSN network, this functionality can be added to the mobile application.

V. TESTS

In order to ensure that the WISE-MUSE mobile application meets all requirements and all features important to the proper monitoring of a museum, it was necessary to carry out usability tests. These tests intended to verify whether such features are implemented correctly.

Tests should be performed with the most appropriate type of users, which, in this case, are the museum's staff. They are the ones who daily deal with the concern of preserving the artworks and, thus, quickly understand the goal of the application, ensuring that it is used and tested the appropriately.

This staff was contacted and visited to conduct the tests. After the tests were carried out, they were asked to answer a short survey to help in better understanding the results of this experience and, thus the usability of the developed mobile application.

The usability tests that were conducted allowed validating the different functionalities of the application, permitting to observe how the user reacts, interacts and navigates through different application scenarios. Also, these tests allowed understanding what the users' preferences are. In addition, these tests help to observe the

 ISBN: 978-1-61208-296-7

performance of users, enabling the analysis of the product's usefulness and ease of use, which allows improving the quality of the developed application.

The tests were conducted with five heads of different museums, aged between 35 and 60. This application has also been tested in the Whale Museum.

It was observed that most users easily interact with the application using all functionalities correctly, especially those who are familiar with an Android phone.

Nevertheless, the tests helped in detecting which aspects to be refined in order to improve the operation for this application, namely:

- To add the ability of zooming in and out to improve the visualization of the historic;
- To provide real-time gradient maps of the measured parameters; and;
- To increase the visibility of notifications in the users' mobile phone, to ensure that abnormal situations are quickly detected.

VI. CONCLUSION

The WISE MUSE project aims to the deployment of a WSN for automatically and continuously monitoring and control the environment of museums. The use of WSNs for environmental monitoring of a museum is, indeed, a more reliable solution. It is also less expensive than manual data collection or than a wired central monitoring system.

This project has proposed several contributions, among them the WISE-MUSE mobile application. This application allows the environmental monitoring based on wireless sensor networks via a mobile phone. Besides, it allows users to constantly check and analyze data on the monitored parameters, anytime and anywhere, enabling monitoring with more flexibility and more security, providing better visualization of all data captured by different sensors through charts, maps and historical gradients, in real-time.

One of the most important features of the proposed application is to provide a notifications' system to users, in case any parameter presents an abnormal value.

The tests carried out allow to conclude that the WISE-MUSE application fits the most important requirements of a museum, associating the efficient display of accurate measurement of environmental factors, all captured by sensors, with the ability of sending notifications, allowing preventive action in preserving the history that these works represent.

VII. FUTURE WORK

In order to make the mobile application more efficient and perceptible to the user, we intend to improve it by providing feedback about the user navigation on the application.

Similarly, we also intend to implement the zoom feature to enable enlarging or reducing the size of the table with the historic of data, so that people who want to browse data in a larger size may do so. We also intend to give the user the possibility of sorting historic data by a parameter chosen by the user, i.e., to allow the user to sort data by date, rooms, floors, or sensors, showing data in a descending or ascending order.

It is desirable that the module generating a gradient map of the environmental parameters is implemented, which will allow displaying data from a particular sensor in a more intuitive manner. This makes it possible to give a more specific idea of what is occurring, allowing the user even to check if a particular sensor needs to be moved to another location or replaced by another one.

The tests of this application were based on data collected from a small testbed installed in the laboratory. The WSN was not installed in the museum yet since the equipment for assembling the monitoring devices did not arrive in time. Thus, the application was tested with some data to permit to observe its basic operation. Therefore, one future work intention is to carry out tests with the WSN deployed in the museum.

TABLE 1. COMPARATIVE TABLE OF THE APPLICATION WISE-MUSE WITH THE SECURITY SYSTEM DEVELOPED FOR ANDROID AND A MOBILE APPLICATION TO MONITOR A WSN [12]

PARAMETERS	SECURITY SYSTEM IN ANDROID [12]	Sensor – Mobile application to monitor a WSN [13]	WISE-MUSE
Presentation of data from sensors in real-time	Yes	Yes	Yes
Presentation of data from emergency doors in real-time	No	No	Yes
Submission of data from cold chambers in real-time	No	No	Yes
Generation of graphics	No	No	Yes
Generation of data historic	No	Yes	Yes
Map view	No	No	Yes
Notification of anomalies	Yes (SMS)	Yes (SMS, E-mail)	Yes (Popup)
Presentation of a video of the monitored site	Yes	No	No
Presentation of a image of the monitored site	No	Yes	No
Visualization of all deployed sensors	No	Yes	Yes
Selection of a specific sensor to visualize	No	Yes	Yes

 ISBN: 978-1-61208-296-7

REFERENCES

[1] L. M. Rodríguez Peralta, L. M. P. L. Brito, J. P. B. F. Santos, J. F. F. Santos, C. M. M. Francisco, C. M. A. Sousa, P. M. Moraes, and F. B. Gouveia, "Environmental Monitoring Platform based on a Heterogeneous Wireless Sensor Network", Cyber Journals: Multidisciplinary Journals in Science and Technology, Journal of Selected Areas in Telecommunications (JSAT), October Edition, pp. 26- 38, ISSN: 1925-2676, 2011

[2] L.M. Rodríguez Peralta, L.M.P. Leão Brito, B.A. Teixeira Gouveia, D.J.G. Sousa, and C.S. Alves, "Automatic monitoring and control of museums' environment based on Wireless Sensor Networks", International Journal EJSE – Electronic Journal of Structural Engineering, Special Issue: Wireless Sensor Networks and Practical Applications, pp. 12 – 34, 2010, ISSN 1443-9255.

[3] Spinwave Systems, April 2009, from http://www.spinwavesystems.com/ [Retrieved: July, 2013]

[4] L. M. P. Brito, L. M. Rodríguez Peralta, F. Santos, and R. Fernandez, "Environmental Monitoring of Museums Based on Wireless Sensor Networks", Proc. 4th International Conference on Wireless and Mobile Communications (ICWMC 2008), IEEE Computer Society Press, Athens, Greece, July 27 - August 1, 2008, pp. 364- 369.

[5] L.M. Rodríguez Peralta, L.M.P. Leão Brito, and J.F.F. Santos, "Improving Users' Manipulation and Control on WSNs through Collaborative Sessions", International Journal of Knowledge and Web Intelligence (IJKWI), Special Issue on Social Media Support for Intelligent Service and Interaction, Vol. 3, No. 3, 2012, pp. 287 – 311, DOI: 10.1504/IJKWI.2012.050860.

[6] MySQL, form http://www.mysql.com/ [Retrived: July, 2013]

[7] Wikibooks, "Structured Query Language", from http://en.wikibooks.org/wiki/Structured_Query_Language/ [Retrived: July, 2013]

[8] F. Rahman, (2011) "Connection between PHP (server) and Android (Client) using HTTP and JSON", from http://fahmirahman.wordpress.com/2011/04/21/connectionbetween-php-server-and-android-client-using-http-and-json/ Retrived: July, 2013]

[9] Overseer , Overseer Network Monitoring Software, Environmental Monitoring Software, 2012, from http://www.overseer-network-monitor.com/Sensatronics-Environmental-Temperature-Monitoring-Software.aspx, [Retrieved: July, 2013]

[10] SensaTronics, "Model EM1 Environmental Monitor ", from http://www.sensatronics.com/wp-content/uploads/2013/07/SensatronicsDatasheets5-4.pdf [Retrived: July, 2013]

[11] Envirosense, EnviroSense Strategic Environmental Consultants, 2012, from http://www.envirosense.com/index.php?, [Retrieved: July, 2013]

[12] P. Heming, J. Linying, Y. Liu, and Y. Kun, "Design and Implementation of Android Phone Surveillance System", Proc. of the IEEE International Forum on Information Technology and Applications (IFITA'10), IEEE, 2012, pp. 222-225.

[13] N. Moreira, M. Venda, C. Silva, L. Marcelino, and A. Pereira, "@Sensor – Mobile Application to Monitor a WSN", 6th Iberian Conference on Information Systems and Technologies (CISTI), Leiria, Portugal, July. 2011, pp. 1 -6.

[14] H. Pang, L. Jiang, L. Yang, and K. Yue, "Research of Android Smart Phone Surveillance System", International Conference on Computer Design and Applications (iCCDA), Shenyang, China, vol. 2, June. 2010, pp. 373 – 376.

 ISBN: 978-1-61208-296-7

Smart Shopping Cart for Automated Billing Purpose using Wireless Sensor Networks

Udita Gangwal, Sanchita Roy, Jyotsna Bapat
International Institute of Information Technology - Bangalore
Bangalore, India
e-mail: udita.gangwal@iiitb.org, sanchita.roy@iiitb.org, jbapat.iiitb@gmail.com

***Abstract*—With the increasing employment of broad area Wireless Sensor Networks (WSN) in the field of consumer applications, it becomes imperative to address the concerns raised by its application, such as reliability, energy consumption and cost-effectiveness. In this paper, we describe the implementation of a reliable, fair and cost efficient Smart Shopping Cart using Wireless Sensor Networks. Such a system is suitable for use in places such as supermarkets, where it can help in reducing man power and in creating a better shopping experience for its customers. Instead of making the customers wait in a long queue for checking-out their shopped items, the system helps in automating the billing process. Along with this ability, the system design also ensures detection of cases of deception invoked by dishonest customers, which makes the smart system fair and attractive to both the buyers and sellers. The system design along with the experimental setup are presented. The results are encouraging and with the use of repeaters at appropriate locations inside the supermarkets, our approach illustrates itself to be conceivable for use outside the laboratory, in real world deployment.**

***Keywords*-Wireless Sensor Networks; broad area WSN; Smart Shopping Cart; Load cell; Image comparison algorithm**

I. INTRODUCTION

Enormous amount of advancements in the field of Wireless Communication has given way to several new technologies and fields altogether. One such upcoming field is Wireless Sensor Networks (WSN), which is maturing at a very fast pace because of its suitability in a wide range of application areas. It consists of a large number of small, low-power, cost-effective, autonomous devices termed as sensor motes. When interfaced with sensors and actuators, which could be simple or complex, they play the combined role of environment-sensing, special-computing and wirelessly communicating devices. These factors accompanied by the effectiveness of technologies for miniaturization of hardware (microcontrollers and radio modems, for example), technologies for sensing equipments, technologies for energy saving and scavenging, and the fact that many applications cannot be wired, makes it suitable for various application domains. Examples of such applications are medicine and health care, disaster relief applications, environment and industrial monitoring, etc. [1]

In this new era of consumerism, broad area WSN finds its use in consumer application areas such as Smart Home, Smart Grid, etc. The challenges here are to not only make the system intelligent by automation, but also to handle the concerns that are raised due to the automation process such as probability of false alarms, energy consumption, cost-effectiveness, etc. Since many sensor nodes are required over a broad area for environment-sensing, the system design needs to concentrate on aspects such as the choice and placement of sensors within the area, communication among the various nodes so that it works reliably with minimum energy requirement and be cost-effective at the same time. In this work, we take the particular case of supermarkets, where our design based on WSN is used to address the following issues:

1) Customer dis-satisfaction because of long waiting time for check-out process, and
2) Involvement of a lot of man-power, which is expensive.

In order to achieve this, we have come up with a design that automates the billing procedure and saves the customers' time. Automation has its own problems. Absence of human operators can potentially lead to inconvenience when the underlying technology fails. It can also lead to dishonest behavior of the customers. We propose and implement a solution that has redundancy built into it in order to reduce the probability of failure, and has three main benefits:

1) It creates a better shopping experience for the customers by saving their time.
2) It minimizes the man-power required at the shopping mall, as the checking-out process at the check-out counters is eliminated altogether.
3) It handles cases of deception if any, thereby making the system attractive not only to the customers, but also to the sellers.

A number of attempts have been made to design a Smart Shopping Cart with various different functionalities. Awati and Awati [2], describe a Smart Trolley design that concentrates on how to get the customers rid of dragging heavy trolleys and to automate billing, but it assumes all the customers to be honest and hence does not tackle cases of deception, if there are any. Further, Yew *et al.* [3] propose a smart shopping for future where the barcodes are completely replaced by Radio Frequency Identification (RFID) tags and scanners. This idea might take a long time to be deployed as it is expensive both in terms of money and energy. A lot of other works describe how products in a store could be tracked by customers instead of spending a lot of time searching for it.

In this paper, the system design considerably minimizes the overhead of wireless communication among the devices involved in the system as almost every processing is done locally at each cart instead of transmitting packets to another node. Hence even when there are a lot of customers present in the shopping mall, there will not be any deterioration in the performance owing to communication gridlock. Every Shopping Cart is equipped with a sensor mote, a load-cell fitted at the base of the trolley, a camera fitted on the top (also acts as barcode scanner) and a system for local processing and display purposes as shown in Figure 1.

Every customer is identified by the ID of the cart s/he picks for shopping. The Base Station at the payment counter consists of a database that stores information of all the products, and a sensor mote to communicate with all the Smart Carts in the mall. When a customer starts shopping, s/he has to scan the barcode of the product with the barcode scanner present at the cart, after which the product has to be put into the basket. The barcode of the product is wirelessly transmitted by the mote to the Base Station using the IEEE 802.15.4 (ZigBee Protocol) [4] over the ZigBee network. ZigBee is chosen

 ISBN: 978-1-61208-296-7

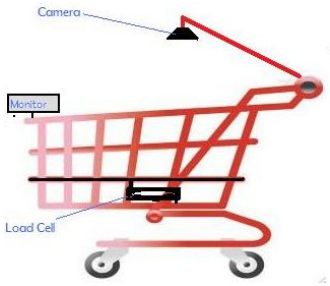


Fig. 1. Smart Shopping Cart

along with the IEEE 802.15.4 compatible sensor motes because they are easily available and mass produced. However, any other short distance radio system will work equally well. In reply, the Base Station sends relevant information about the product, which is used in the decision-making process at the cart. In order to handle all the cases of mistake/dishonesty, the design includes the use of image processing at the cart. After the customer finishes shopping, s/he then proceeds to the payment counter to pay the bill amount and is assisted by an attendant only in the case the system detects discrepancy in the self check-out process of the customer.

The organization of this paper is as follows: Section II presents the detailed system design, Section III gives the implementation details, Section IV discusses the result and feasibility issues, and Section V concludes the paper.

II. Detailed description of the Smart System

The features supported by the Smart Shopping Cart and the idea behind how these features are achieved are explained in the next two sub-sections.

A. Features of the Smart Shopping Cart

The capabilities of the Smart Shopping Cart are listed below:

1) The basic function of calculating and updating customers' bill as and when s/he places the shopped products in the cart.
2) The customer can also track the details of the purchased items as well as the current bill amount on the monitor that is attached to the cart.
3) In addition to the above features, it also includes the handling of the following special cases, which ensures that the system is fair in all respects. All the cases mentioned below are detected by the system.
    a) Attempt to take away products by keeping them into the cart without scanning their barcodes.
    b) When the customer scans a product, but forgets to keep it in the cart.
    c) Attempt to scan one product, but place multiple products in the cart.
    d) Attempt to take away one product of higher price by scanning the barcode of another product of lesser price.
    e) Since consumers are likely to change their mind, our implementation allows for removing any item already placed in the cart, without help from attendant.

Next few sub-sections describe how these functionalities are incorporated into the Smart Shopping System.

B. The Design Idea

The design has been focused to tackle all the scenarios which are mentioned above. As the goal of the Smart Shopping System is automation, the first requirement is to have a barcode scanner attached to every shopping cart. Hence, this design includes a *camera-based barcode scanner*, which is fitted to the cart. The barcode scanner is required to identify a product so that its price can be determined from the database, which stores all the relevant information about all the products. The database in our design is stored in the Base Station, which is located at the payment counter. Some of the information per product that is stored in the database includes its barcode, its name, price and weight.

The *weight* attribute of a product has been chosen for a way to double-check the identity of the product in order to detect deception in the system. A *load-cell* has been configured as a weight sensor. The output of the load-cell is used in the decision making process at the cart. If the weight of a product estimated by the load-cell is not the same as the actual weight of the product, it is interpreted as a case of discrepancy.

The design involves a third level of check to further enhance the decision-making process, which makes use of *Image Processing*. While the barcode of the product is being scanned, a picture of the product is taken by the same camera that also works as the barcode scanner. If a person wants to exchange this product with a costlier one, it will be after scanning the barcode that he will do so. There is a slab attached to the top of the cart which is meant to play the role of placing the products into the cart when it is triggered to do so, instead of the customer having to put the product into the cart by himself. The person places the product on this slab once the scanning is over. Another picture of the product is taken just before the slab lets the product into the cart. Both the images are stored locally in the system present at the cart. An *image comparison algorithm* is run on these two images to find if they are the same products. If they are not found to be the same, it is interpreted as a case of discrepancy. The two images are removed from the memory of the system just after obtaining the result from the algorithm in order to restrict the memory usage of the systems at the cart.

The processing is done locally instead of transmitting the image for every product to the Base Station for comparison, in order to reduce the overhead on the wireless communication, which makes it energy-efficient. This also ensures that the system gives the same performance even with a lot of customers in the store shopping at the same time.

C. Operation of the Smart Shopping System

A customer enters the Smart Shopping Centre. On entering, s/he first picks a Smart Shopping Trolley. Each trolley is given a unique ID and every customer is associated with the ID of the trolley chosen. A typical trolley is expected to look like the one shown in Figure 1. The functioning of the system is listed below:

- When the customer picks up a product that s/he wishes to purchase, s/he first scans the barcode of the product using the barcode scanner and then places it on the slab of the cart, which is meant to play the role of putting the products into the cart when it is triggered to do so. While the customer is scanning the barcode of the product, a picture of the product is taken and stored in the system's memory. The barcode and the cart ID are transmitted as two different fields in a single Zigbee packet by the sensor mote on the cart to the Base Station. An additional field called the *attendant-flag* field is sent only in case of discrepancy.
- At the Base Station, this transmitted information is received by the sensor mote attached to it. This information is then used to fetch relevant information about the product from the database corresponding to the barcode. The database consists of the following details at least: the barcode, name of the product, price and weight. The weight and price fields corresponding to the received barcode are extracted and kept aside.
- Meanwhile, at the cart, the slab still holds the product and another picture of the product is taken *just before* the slab lets the product into the trolley. An image comparison algorithm is run once it has both the images. Depending on whether the

 ISBN: 978-1-61208-296-7

images match or not, it sets the *attendant-flag* field, which is later transmitted to the Base Station for it to take appropriate actions.

- Once the product is inside the trolley, the role of the load-cell comes into play. The weight of the product is estimated and then transmitted to the Base Station using the same mote on the Smart Cart.
- At the Base Station, the weight which is received from this cart is compared with the weight that was retrieved earlier from the database corresponding to the same cart ID. Depending on whether the weight matches or not, appropriate actions are taken.
- This procedure is repeated for every product the customer purchases. Finally, when the customer finishes shopping, s/he goes to the counter in order to pay the bill amount. In case of any detected discrepancy, an attendant verifies the self-checkout process carried out by the customer. The attendant is signaled by the User Interface present in the Base Station.
- If the two weights are found to be equal *and* if the Base Station does not receive the attendant-flag field, then on entering the customer's cart ID on a particular field in the user interface, it displays the detailed bill of the customer's purchase along with a green symbol. This implies that the customer can pay the bill amount and carry on.
  On the other hand, if the attendant-flag field is detected *or* if the two weights are found to be different at the Base Station, then on entering the customer's cart ID, it displays a red symbol and an alarm sets off, indicating that an attendant has to request the customer to wait for the check-out process again.

If a customer changes his mind, the reverse process has to be carried out. After the customer takes the product out, it has to be scanned and the image of the scanned product is then captured. The Base Station has been programmed to handle this case which enables the customer to do so. This implementation also takes care of all discrepencies in the same manner as described above.

III. Implementation Details

A prototype has been made based on the the same design idea. The various components that are used in the implementation along with the important considerations are explained in details.

1) Barcode Scanner: The prototype uses a camera-based barcode scanner for implementation, which uses a small video camera to capture an image of the barcode and then use sophisticated Image Processing techniques to decode the barcode. We have used a webcam for this purpose, which is supposed to be fixed at the top, facing the slab attached to the cart. The ZBar barcode reader [5] is used for the implementation, which supports many popular symbologies (types of barcodes). It is made to run on the Linux (Ubuntu) Operating System. It also has a user interface that is displayed on the monitor in which the customer can see the green lines along the barcode if it has been detected correctly or a red light if it has not been detected. Figure 2 shows the two conditions.

Fig. 2. User interface of the barcode scanner showing the detection and non-detection of barcode

2) Weight Sensor: A load-cell is configured as a weight sensor. A load cell is a transducer, which is used to convert a force into electrical signal, an analog output voltage. The load cell CZL601-3kg [6] shown in Figure 3 has been used for the experiment, where *3kg* denotes the Rated Capacity of the load cell. The load cell can be chosen based on what precision in weight is required, which in turn depends on what kinds of products are available in the Shopping Store. The cost of the load-cell depends on its precision; higher the precision, higher the price.

One end of the load cell has to be fixed and force has to be applied on the other end so that the deformation in the strain gauge of the load cell is indirectly converted to an output voltage. The load cell is supplied with a DC voltage of 9 Volts with the help of a Transistor-battery.

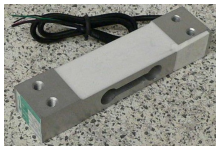


Fig. 3. Load cell CZL601-3kg

TABLE I
SPECIFICATIONS OF INTEREST FOR CZL601-3KG

Rated capacity	3 kg
Rated output	1.948 mV/V
Excitation voltage provided	9 Volts

Rated capacity is the maximum axial load that the load cell is designed to measure within its specifications. The maximum output voltage that can be provided by this load cell is 1.948 mV $\times$ 9V = 17.532 mV. The load cell gives an output voltage which is *almost* proportional to the weight that is applied. It is not exactly linearly proportional to the weight due to many factors such as hysteresis error, repeatability and temperature effects. Readings have been taken as shown in Figure 4, in order to calibrate the output voltage of the load cell for a given input weight. The straight black line shows the ideal response, whereas the blue line shows the actual response. The response flattens out towards the rated capacity.

The output voltage of the load cell is typically in the order of a few millivolts and requires amplification before it can be used. A μA741CN operational amplifier has been used as a negative feedback non-inverting amplifier. An amplification factor of 11 is achieved by the arrangement shown in the Figure 5.

This amplified output is then given to an ADC on the sensorboard. The output of the ADC varies with the amplified output of the load cell, which in turn varies with its input weight. Each cart stores a

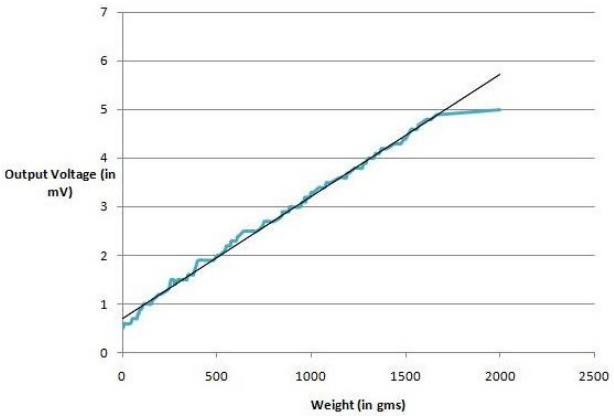


Fig. 4. Plot for Weight versus Output Voltage

 ISBN: 978-1-61208-296-7

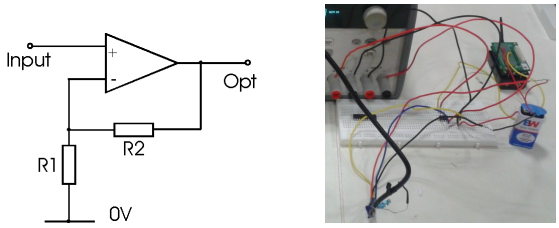


Fig. 5. Voltage amplifier circuit

look-up table, which consists of a mapping between the ADC output values and its corresponding input weight. This is required to compare the weight of the product that has been actually put into the basket with the weight field that the Base Station retrieves from the database. Also, it is advantageous if every cart has its own mapping for the different weight values because load-cells with different precisions could be used for different carts. This effectively means that for different sections of the shopping mall, carts with different precision of the load-cells could be used, instead of using a high precision load-cell throughout, which makes it more cost-efficient.

3) Image Comparison algorithm: The Image comparison algorithm that is chosen for the design is the SIFT (Scale-Invariant Feature Transform) algorithm. It extracts interesting points on the object in the image to provide a *feature description* of the object. These features extracted from the training image are then used to identify the object when attempting to locate this object in a test image containing many other objects as well. This algorithm is apt for the design as the algorithm works even when the object in the two images are not same in size, orientation and scale. So, if the customer places the product on the slab even in a rotated direction, it still identifies whether the product is the same or no. The algorithm works well for all lighting conditions except for very extreme lighting condition, i.e., with almost no light in the ambience.

4) Sensor mote at the Smart Cart: All the Smart Carts are equipped with a Crossbow IRIS-XM210 [7] mote running TinyOS [8] Operating System, along with a MDA100CB sensor board [13]. The mote is connected to the system which is present at the cart via USB cable for monitoring and display purposes. The system at the cart is programed to calculate the weight (with the help of the look-up table) two seconds after the product is being sensed by the load cell so that even if the product is dropped with a great force into the cart, it does not estimate a wrong weight based on the initial momentary thrust on the load-cell. The same sensor mote is used to transmit all the information to the Base Station - the barcode, the trolley ID and the flag fields.

5) The Base Station: The Base Station resides at the counter meant for payment for bills. A PC with an IRIS sensor mote is used to communicate with the Shopping Carts. The database which contains the information of all the products that are present in the store, resides in the PC. MySql database has been used for the implementation. It consists of a table which consists of the following fields: (i) Barcode ID (Primary key), (ii) Name of the product, (iii) Price, and (iv) Weight of the product. The PC also supports a GUI meant to assist the customers to pay their bill amounts and to alert the attendant in cases of discrepancy. Figure 6 and Figure 7 show the laboratory set-up of the proposed system. Experiments have been conducted using the set-up and various products have been selected for testing purpose. Figure 8 and Figure 9 show the UI at the Base Station, in which the red and the green light is indicative of whether the system has detected a case of discrepancy or not respectively.

IV. Result and Feasibility

The experimental set-up is tested for various test cases, with various products tested for all the possible cases mentioned in

Fig. 6. Prototype model of a Smart Shopping Cart

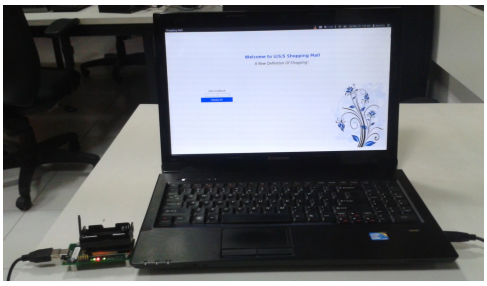


Fig. 7. Prototype model of Base Station

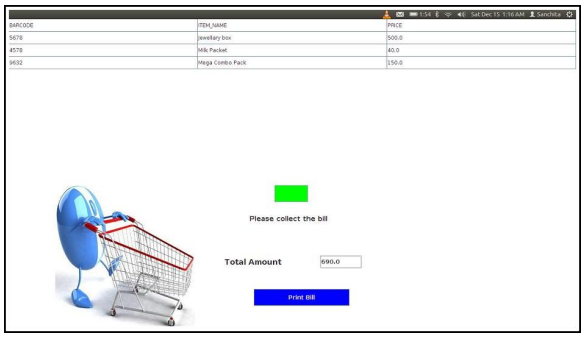


Fig. 8. Generated bill with the Green indicator

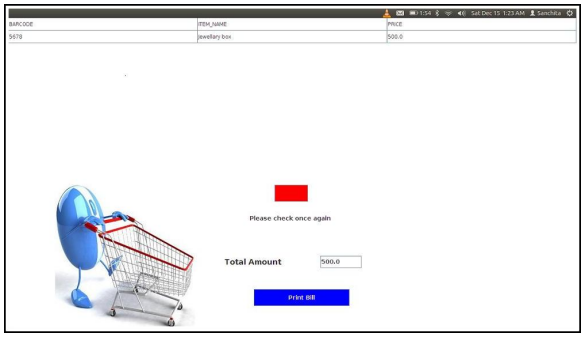


Fig. 9. Generated bill with the Red indicator

 ISBN: 978-1-61208-296-7

Section II. When the system is tested with a single Shopping Cart and a Base Station, it gives the correct result for all the cases except for the case when the lighting condition is very poor, i.e., when the lighting condition in the environment is very dim/dark. This is because the object in the image cannot be recognized because of the darkness, due to which the SIFT algorithm fails to extract the key points of the object. The lighting in a store is expected to be bright. Low lighting conditions can be indicated on the smart cart by setting the attendant flag. This attendant flag is the same as the one set when weight or images do not match.

Next, we observe how much time it takes for the entire process to take place with respect to the distance of a Shopping Cart from a Base Station. This is required in order to decide on the placement and the number of repeaters inside the Shopping Mall. The processing time includes the time taken by the cart to generate a decision and the time for the wireless communication between the Base Station and the Shopping Cart. Figure 10 shows a plot of the processing time against the distance of the cart from the Base Station. This variation in the response time is mainly due to the time taken for the wireless communication, as the time taken in decision-making at the cart is approximately the same every time.

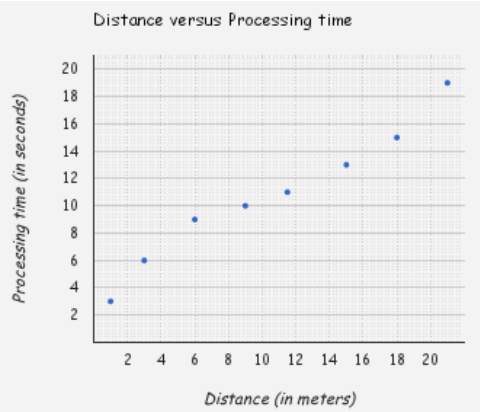


Fig. 10. Distance versus Processing time

The range for mote-to-mote communication is found to be 21 meters inside a busy building at our Campus. This brings in the need for wireless repeaters depending on the dimensions of the Shopping Mall. According to the information gathered, the typical area of a Shopping Mall is about 2,500 square meters per floor [14]. If we consider the layout of such a store to be a square (for a rough estimation), the length per side turns out to be approximately 50 meters. This implies that if the system has to be deployed in such a place, there has to be few repeaters located at different positions in order to ensure coverage of the entire broad area. The IRIS motes can act as repeaters too if they are programed to do so. If there are 2 payment counters (Base Stations) per floor on the opposite corners, 2 to 3 repeaters per floor will suffice.

V. Conclusion and Future Work

The project successfully demonstrated the possibility of using WSN for developing a Smart Shopping System which automates the entire billing procedure. The system which is developed is highly reliable, fair and cost-effective. It is reliable and fair because of the effectiveness of WSN combined with a highly reliable Image Processing technique. The system is also energy constraint as it uses a passive sensor and it reduces the communication requirement. The decision making process is done locally within the cart, thereby eliminating an overhead to the communication between the motes. Also, the application does not make use of complex routing mechanisms or unicast transmissions; our implementation makes use of the simple broadcast technique to communicate with the Base Station as each cart is associated with a unique ID. The system is cost-effective as it requires only one passive sensor (the load-cell) and a camera-based barcode scanner (which is way cheaper than any other type of barcode scanners) per cart. In the bigger picture, it reduces the man-power requirements.

The effect of multiple users operating at the same time, as well as any spectrum coexistence issues must be studied since the proposed system uses the over-used 2.4 GHz spectrum. The current implementation also does not talk about the placement of repeaters inside a supermarket layout.

References

[1] H. Karl and A. Willig, “Protocols and Architectures for Wireless Sensor Networks,” Chichester, England, 2005.
[2] J. Awati and S. Awati, “Smart Trolley in Mega Mall,” vol. 2, Mar 2012.
[3] L. Yew, L. Fang, C. Guancheng, C. Jianing, and L. Hangzhi, “RFID: Smart Shopping for the future,” Singapore Management University, Tech. Rep.
[4] Ergen, S. C., “ZigBee/IEEE 802.15.4 Summary,” EECS Berkely, September 2004.
[5] “ZBar bar code reader,” http://zbar.sourceforge.net, [retrieved: July 2, 2013].
[6] “CZL-601, 3-120 Kg [Aluminium load cell],” http://www.saithongelectric.com/index.php?lay=show&ac=article&Id=518670&Ntype=22, [retrieved: June 28, 2013].
[7] “Data Sheet for Memsic IRIS Motes,” http://www.memsic.com/products/wireless-sensor-networks/wireless-modules.html, [retrieved: July 15, 2013].
[8] D. Gay, P. Levis, and D. Culler, “Software Design Patterns for TinyOS.”
[9] D. Gay and P. Levis, “The nesC Language: A Holistic Approach to Networked Embedded Systems.”
[10] “Tiny OS Documentation Wiki,” http://docs.tinyos.net/, [retrieved: July 12, 2013].
[11] “Tiny OS 2.0.2 Documentation,” http://www.tinyos.net/tinyos-2.x/doc/, [retrieved: July 12, 2013].
[12] D. Gay, P. Levis, and D. Culler, “nesC 1.1 Language Reference Manual.”
[13] “MTS/MDA Sensor Board Users Manual,” Crossbow Technology, Inc., June 2007.
[14] “India’s Largest Malls - 2010,” asipac, Mar 2010, [Research Studies On Malls in India].

 ISBN: 978-1-61208-296-7

Activity Recognition Using Wearable Sensors for Healthcare

Annapurna Soumya Evani*, Bharadwaj Sreenivasan†, Joshi Shruti Sudesh‡, Monika Prakash§, Jyotsna Bapat¶
International Institute Of Information Technology, Bangalore, India
{annapurna.soumya*,bharadwaj.s†,Shruti.Joshi‡, Monika.P§}@iiitb.org
jbapat@iiitb.ac.in¶

***Abstract*—Wireless Sensor Networks (WSNs) formed by wireless sensors can be designed to interact continuously with the real world. WSNs help in communicating accurate real time information and have proliferated from the areas of industrial processing and environmental monitoring to science and healthcare. Today's healthcare systems are facing two serious challenges: rapid growth in adult population and severe nursing shortage. As per United Nations (UN) statistics, the number of people in the world aged 60 years and above are expected to increase from 605 million to 2 billion by 2050. Nursing shortage has been identified as a global crisis since 2002 and India has a nurse density per thousand population of 0.80. These statistics suggest a urgent need for automated monitoring systems to aid the healthcare industry. In this paper, we describe the implementation of a patient activity monitoring system using wearable flex sensors. Patient activities are classified as standing, sitting, or walking using a lightweight, low latency algorithm. Our approach features a unique sensing technique and facilitates a cost effective and energy efficient solution for healthcare.**

Keywords*—Healthcare; activity recognition; flex sensor; wireless sensor networks.

I. Introduction

A. Motivation

The increase in the number of people availing healthcare services has stressed the need for a larger nursing pool. According to global statistics, the demand for registered nurses is expected to grow from 2 million to 3.2 million between 2008 and 2018, a 60% increase [1]. In the Indian healthcare sector, there is a shortage of 350,000 nurses, with a nurse to patient ratio of 1:1205 [2]. As per World Health Organization (WHO), India is ranked 52 out of 57 countries facing Human Resources for Health (HRH) crisis [3]. With this global shortage, patient care is expected to be compromised. There have been reports of nurses being stressed and overworked resulting in significant attrition. In such a grave scenario, it is necessary to think of alternatives for taking care of the elderly population. One approach is to reduce the need for around-the-clock care givers and automate the monitoring processes in the healthcare industry.

B. Background

Applications of WSNs (Wireless Sensor Networks) in healthcare has been an area of interest in recent years. These applications require collecting large data sets from multiple sensors and manipulating the context to find the patterns pertaining to the medical aspect being monitored. The parameters such as security, privacy, user friendliness, scalability, context-awareness are critical in designing healthcare monitoring application [4]. There have been significant contributions in the area of vital sign monitoring in hospitals by providing assistance to patients with declining sensory and motor capabilities, at-home assistance to patients suffering from Alzheimer's disease, and depression for the elderly population and many more [5]. The sensors in smart health systems are capable of sensing heart activity, temperature, brain activity, glucose levels, behavioral patterns, etc. [6]. Researchers in computer, networking, and medical fields are working together to create intelligent healthcare monitoring systems. In the future, personal area network technologies such as radio frequency identification (RFID), Bluetooth, ZigBee, and wireless sensor networks are expected to work together with infrastructure based networks to provide context-aware applications [4].

In this paper, we propose an approach in-line with one of the current on-going research interests and design principles, i.e., activity detection for patients. Activities such as sitting, standing, and walking are recognized by the system using wearable sensor. This activity data can be further analyzed to know whether the patient is healthy and is following his/her routine day-to-day activities without constant vigilance by a real person. Also, it is possible to detect inactivity, which can be treated as an abnormal behavior and an alarm can be triggered alerting concerned people.

C. State of the Art

Existing activity recognition systems can be broadly classified into video sensor-based activity recognition and physical sensor based activity recognition [7]. Video sensor-based activity recognition is a traditional method, which was extensively used [8][9][10][11]. It involves continuous capture of human activities through cameras; however, the video method raises privacy issues and requires the individual being monitored to either remain within the vicinity of the camera or to have a video recorder attached onto the body [12]. Moreover, the extraction of features from the captured images requires complex computations. Due to these limitations, wearable sensor based activity recognition, which requires less data processing (physical sensor based activity recognition) is recommended. Inertial sensors - accelerometer and gyroscope are the most widely used body worn sensors for activity recognition [12][13][14]. Although these sensors are accurate in monitoring activities, the proposed solutions using them are complex and expensive.

To overcome these challenges, we propose a novel method of human activity recognition using a low weight wearable flex

 ISBN: 978-1-61208-296-7

sensor. This low cost high resistance sensor enables the product to be an affordable solution in detecting patient activities in hospitals.

This paper is organized as follows: Section II discusses System Architecture comprising System Design and Operation. Section III describes Implementation Details explaining the Algorithm of the system. Section IV explains the Experimental Setup and Results. Sections V and VI discusses the conclusion and future scope, respectively.

II. System Architecture

A. Overview

The posture recognition system consists of three main components: the Flex sensor, the Controller, and the Wireless Transceiver. The flex sensor is placed on the kneecap of the patient. Based on readings from the flex sensor, the controller can make intelligent decisions about the posture of the patient. This information is transferred to a central base station, which collates the results from different controllers. This allows a centralized database to be created for all patients using the posture recognition system. The doctor or nurse has the freedom of monitoring several patients with the help of a centralized architecture. The patients can conduct their day-to-day activities without being deprived of their privacy.

B. System Design

The Project uses an IRIS - XM210 mote with an MDA100CB sensor board. The Sensor board hosts the flex sensor as shown in Figure 1.The flex Sensor is a resistive device used to detect the flexing of the entity hosting it. The

Fig. 1. Sensor with the sensor mote

specifications of the sensor are shown in Table I.

TABLE I. Flex Sensor Specifications

Manufacturer	Spectra Symbol
Length (inches)	4.5
Flat Resistance	10KΩ
Resistance Tolerance	± 30%
Bend Resistance Range	60KΩ - 110KΩ
Placement	Knee

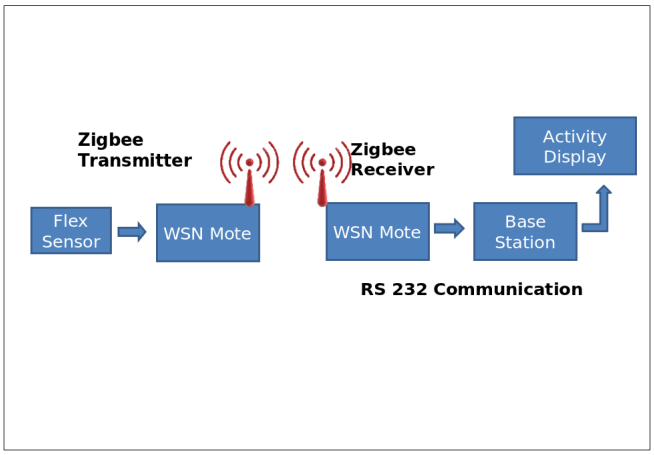


Fig. 2. System Architecture

C. System Operation

The block diagram of the system is shown in Figure 2. The operation of the system can be broadly classified as a three phase phenomenon - Sensor data collection, In-network processing and Activity display.

Sensor data collection

The flex sensor's resistance increases linearly with the bend of the sensor. In order to measure the voltage across the sensor, a voltage divider circuit is constructed as shown in Figure 3.

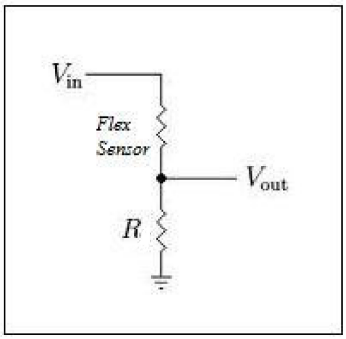


Fig. 3. Flex sensor circuit

$$V_{out} = \frac{R * V_{in}}{R + R_{flex}} \quad (1)$$

The output V_{out} is given by equation (1). The 10 bit Analog to Digital Converter (ADC) present in the IRIS mote is programmed to sample the raw sensor data (V_{out}) at regular intervals. The digitized sample values are considered for further processing in the sensor node.

In-network processing

This phase is the most important phase in sensor networks. To minimize the energy expenditure, radio transmissions must be reduced by local processing of the data. The sampled values are processed using a threshold technique to determine the activity of the patient. The number of transmissions is reduced

 ISBN: 978-1-61208-296-7

by detecting only the change in the posture of the patient, which is then communicated. In order to know whether a node is alive, the posture of the person is sent to the base station with a longer periodic interval, regardless of the change in state of the person.

Activity Display

The base station is a ZigBee-compliant device, which operates at 2.4 GHz range. The base station sends the received packets to the central coordinator over a serial link. The central coordinator hosts the application, which interprets information from the received packets and displays the current activity of the patient in a GUI. Each packet is tagged with a 2-byte Mote ID, which is unique for a given sensor mote. By using the Mote ID, the base station differentiates data from different motes and generates activity history of every patient.

III. Implementation Details

The posture classification algorithm primarily classifies the states into a static or a dynamic state. Sitting and standing falls into the category of a static state, whereas walking and activity transitions are part of the dynamic state. The algorithm is maintained asymptotically less complex, as energy consumed for processing is lower compared to transmitting.

Another important parameter considered during the implementation of the algorithm is response time. The finite size of the memory used in storing the samples is a constraint in achieving the required accuracy in real time applications; however, the larger the memory, greater the delay incurred in processing. To maintain this trade-off between accuracy and delay, optimal memory size is chosen.

The flowchart of the system is shown in Figure 4. The

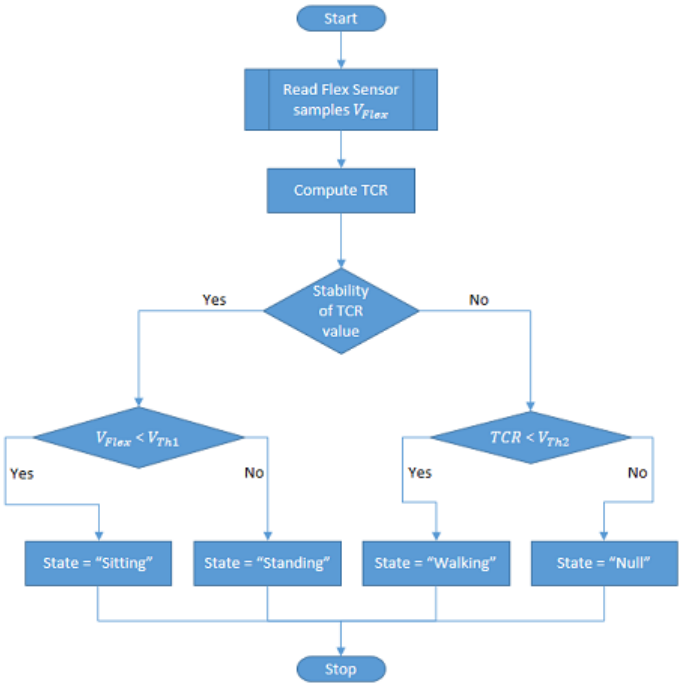


Fig. 4. Flow Diagram at Sensor Mote

postures can be subdivided into two categories: static and dynamic. To classify postures as either static or dynamic, we have created a metric, the Threshold Crossing Rate (TCR). The TCR is computed based on the rate of oscillation of voltage signal V_{flex}, where V_{flex} is voltage measured across the series resistor R as shown in Figure 3. The TCR is further processed to determine between static and dynamic states. It should be mentioned that the walking speed of the patient does affect calculation of TCR. After the determination, we further process the TCR signal to find out whether it is one among the following states: sitting, standing, or walking.

Static postures such as sitting and standing are prone to noise in the V_{flex}. This noise is due to the sensitivity of V_{flex} to the movements of the patient. In order to remove such jitter, we compute the stability of V_{flex} for a window's worth of data. After finding whether the signal is stable, we use a simple thresholding technique to determine between the sitting and standing states. Transitions between activities can cause V_{flex} to look similar to a walking scenario. This introduces false states into the decision system and thus can reduce the accuracy. To remove such transitions, a fourth state is introduced, which is called the "Null" state. All unknown oscillations and transitions in V_{flex} are categorized into this "Null" state. To accomplish this, we apply a threshold on the TCR signal, which is also computed empirically.

IV. Experimental Setup and Results

A. Threshold Determination

As per the algorithm, a threshold is used for classifying the patient postures. The threshold is determined for all postures by sampling the sensor values at regular intervals. The experiment was conducted on 30 people, and the readings were found to be stable throughout the experimental phase.

The pattern observed for the activities, - sitting, standing, and walking is shown in Figure 5. The voltage signal shows the flex sensor readings with respect to time. The system state signal is encoded as 0, 0.5, and 1, where the values refer to sitting, standing, and walking, respectively.

The threshold remained relatively constant due to the casing supporting the flex sensor. It helped in measuring precise readings from the sensor. It saved extra overhead required for computation of adaptive threshold, which would have been mandatory due to fluctuations in readings. The deployment of the sensor mote for the experiment is as shown in Figures 6 and 7.

Figure 8 and Figure 9 show the display at the base station used for monitoring the patients.

B. Range, Delay, and Accuracy Measurements

The experiments for range, delay, and accuracy calculations were conducted in a hospital-like environment. The test environment had rooms located on both sides of a long passage. The base station was located at the extreme end of the passage. The system yielded good results for both line of sight and multipath conditions. The activity detection of patients in the rooms covering the opposite end of the base station was also successful. The indoor line of sight range was approximately 500 meters. The delay observed over this range was between 2 - 4 seconds.

Position of the sensor also influence the accuracy of the system. The flex sensor should be placed on the middle of

 ISBN: 978-1-61208-296-7

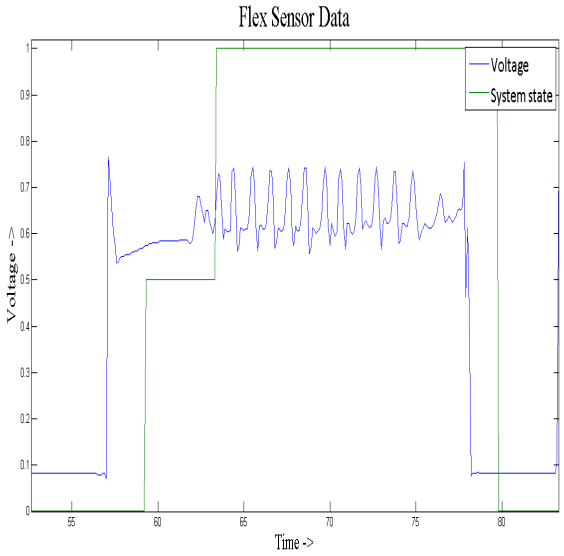


Fig. 5. Threshold Detection Technique

Fig. 6. Sensor view in sitting posture

the person's knee cap so that maximum bend is faced by it. For measuring the success rate, 30 people were asked to perform activities like sitting, standing and walking in a random manner. The percentage of time when the actual posture of the person conformed with decision from the Activity Recognition system was used to compute the success rate. We achieved a success rate of 81% based on the test trails.

V. Conclusion

The project successfully demonstrated the recognition of human activities - sitting, standing, and walking with an accuracy of 81% through a resizable flex sensor band worn on the knee. The developed system is cost effective as only one sensor is required, and also making a flex sensor is feasible through inexpensive materials. Adding to the advantage, the flex sensor is a resistance device that does not require an external power source like active inertial sensors [13]. Also, low energy consumption is maintained by performing in-network processing resulting in long life of the product. Finally, the robust casing and light weight of the product makes it so comfortable that patients would get a feel of just wearing a knee band. Thus, this presents a low cost and a low power solution for remotely monitoring patients in hospitals.

Fig. 7. Sensor View In Standing Posture

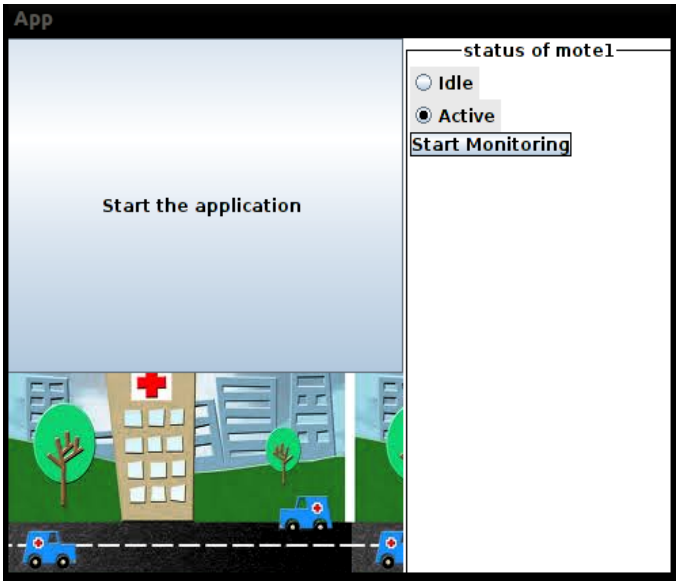


Fig. 8. GUI at the Central Base station

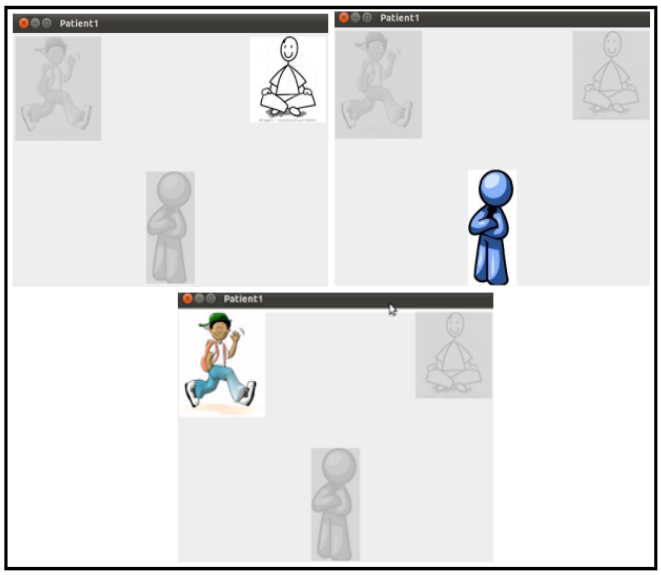


Fig. 9. GUI Depicting Postures

 ISBN: 978-1-61208-296-7

VI. Future Scope

We have proposed a novel activity recognition system using the flex sensor. To make the system cost effective, we propose that the flex sensor be constructed using simple inexpensive materials. The system has the capability of detecting postures such as walking, sitting, and standing. It can be extended to detect the fall and the location of the patient using localization techniques. We also propose that constant human monitoring at the base station can be eliminated by alerting the care giver over SMS.

Acknowledgment

This paper was supported by CEEMS Laboratory at the International Institute of Information Technology, Bangalore funded by the Government of Karnataka. We are grateful to Joseph Jeffrey for his immense support throughout the phase of design and conceptualization of the system.

References

[1] C. A. Holland, N. Ferrell, and D. James. (2009) The nursing shortage: Exploring the situation and solutions. [retrieved: May, 2013]. [Online]. Available: http://www.minoritynurse.com/article/nursing-shortage-exploring-situation-and-solutions

[2] M. Doshi, "Northbridge capital report on hospital sector," November 2010, [retrieved: June, 2013]. [Online]. Available: http://www.northbridgeasia.com/research_reports.aspx

[3] WHO, "World health organisation: World health statistics 2011," 2011, [retrieved: May, 2013]. [Online]. Available: http://uhc-india.org/uploads/ThammaRaoD_HumanResourcesforUniversalHealthCoverage.pdf

[4] H. Alemdar and C. Ersoy, "Wireless sensor networks for healthcare: A survey," *Computer Networks*, vol. 54, no. 15, pp. 2688 – 2710, 2010. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1389128610001398

[5] P. Neves, M. Stachyra, and J. Rodrigues, "Application of Wireless Sensor Networks to Healthcare Promotion," *Journal of Communications Software and Systems (JCOMSS)*, vol. 4, no. 3, 2008.

[6] K. Grgić, D. Žagar, and V. Križanović, "Medical applications of wireless sensor networks - current status and future directions," *Medicinski Glasnik*, vol. 9, 1840-0132. [Online]. Available: http://hdl.handle.net/10400.11/548

[7] A. Jehad Sarkar, D. Guan, T. Ma, W. Yuan, and Y.-K. Lee, "Review of Sensor-based Activity Recognition Systems," vol. 28, no. 5, 2011, pp. 418–433.

[8] X. Xu, J. Tang, X. Zhang, X. Liu, H. Zhang, and Y. Qiu, "Exploring techniques for vision based human activity recognition: Methods, systems, and evaluation," *Sensors*, vol. 13, no. 2, pp. 1635–1650, 2013. [Online]. Available: http://www.mdpi.com/1424-8220/13/2/1635

[9] P. Turaga, R. Chellappa, V. S. Subrahmanian, and O. Udrea, "Machine recognition of human activities: A survey," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 18, no. 11, pp. 1473–1488, Nov.

[10] R. Bodor, B. Jackson, N. Papanikolopoulos, and H. Tracking, "Vision-based human tracking and activity recognition," in *Proc. of the 11th Mediterranean Conf. on Control and Automation*. Kostrzewa Joseph, 2003, pp. 18–20.

[11] C. N. Joseph, S. Kokulakumaran, K. Srijeyanthan, A. Thusyanthan, C. Gunasekara, and C. Gamage, "A framework for whole-body gesture recognition from video feeds," in *Industrial and Information Systems (ICIIS), 2010 International Conference on*, 29 2010-Aug. 1, pp. 430–435.

[12] O. Lara and M. Labrador, "A survey on human activity recognition using wearable sensors," *Communications Surveys Tutorials, IEEE*, vol. PP, no. 99, pp. 1–18, 2012, [retrieved: June, 2013].

[13] H. Junker, O. Amft, P. Lukowicz, and G. Tr "Gesture spotting with body-worn inertial sensors to detect user activities," *Pattern Recognition*, vol. 41, no. 6, pp. 2010 – 2024, 2008, [retrieved: June, 2013]. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0031320307005110

[14] B. Florentino-Liaño, N. O'Mahony, and A. Artés-Rodríguez, "Hierarchical dynamic model for human daily activity recognition." in *BIOSIGNALS*, S. V. Huffel, C. M. B. A. Correia, A. L. N. Fred, and H. Gamboa, Eds. SciTePress, 2012, pp. 61–68, [retrieved: June, 2013]. [Online]. Available: http://dblp.uni-trier.de/db/conf/biostec/biosignals2012.html#Florentino-LianoOA12

 ISBN: 978-1-61208-296-7

Proposed Middleware for Sensor Networks in Cyber-Physical System Environments

Jorge R. Beingolea Garay, Alexandre M. de Oliveira and Sérgio T. Kofuji

Laboratory of Integrated Systems - LSI - Department of Electronic Systems Engineering

Polytechnic School, University of São Paulo

São Paulo, Brazil

{Jorge, amanicoba, kofuji}@pad.lsi.usp.br

Abstract - **This paper describes the modeling proposal of a middleware for sensor networks in Cyber-Physical (CP) environments adopting the reuse of Designs Patterns, which must favor the coupling in the underlying layers of the middleware and facilitate the communication between components in the upper layers. Our main goal is to reach the robustness and flexibility requirements as a result of the high probability of heterogeneity and critical events in Cyber-Physical System (CPS). This modeling is presented for the treatment of events by the middleware, considering functional and non-functional spatiotemporal requirements, as well as the set of requirements common to the CPS. The structure of the middleware, the MVC design pattern and the events model proposed seek to favor the abstraction, processes treatment and events that result from the CPS heterogeneous environment. The result is a lower level of complexity in implementation, simplified control of events and levels control of coupling between the components of the proposed middleware.**

***Keywords**: Middleware; CPS; WSN; Service Design Patterns; Object Design Pattern.*

I. INTRODUCTION

A Cyber-Physical System (CPS) consists of a network of elements that interacts with physical environment and computational tools (sensoring, actuation, control and application). Most of these elements are constituted by embedded hardware devices (sensors and actuators) for monitoring and controlling of physical processes. Generally, the concept of CPS appears as an integration process of computing elements, which has become known by the physical elements and has ceased of being totally dependent. However, their autonomy and functionality will always be conditioned to other systems that extend their activities according to the monitored physical phenomenon and react in a more autonomous way, requiring, in the process, the implementation of other synchronized systems in real time.

The CPS can also be viewed as a management system capable of grouping several applications that function autonomously and include a control loop. However, for this technology to become viable for several contexts in the physical world, it is necessary to consider the limitations of computational elements that integrate this technology. For the case, such limitations become greater when it is a sensor network matter.

Recent researches on CPS showed some of the challenges on building CPS projects, and, among these challenges, we need to investigate new middleware projects for dealing with features such as heterogeneity, components scalability, and critical issues related to sensoring, control, and application. The proposals include the WebMed middleware [1], which is a Service-Oriented Architecture for CPS. On this paper, we highlight the need for more flexible and reusable middleware, taking advantage of the innovative features of web technologies. According to the heterogeneous characteristics of CPS, and their high tendency to scalability, it is easy to predict the excessive source of performance overload that will exist inside of these types of systems. In [2], a Dynamic Control Middleware (DCM) is proposed on the possibility of integration between different CPS. The proposal is limited to the control of IPV6 devices [2].

In many distributed systems, processes communicate with each other primarily through the propagation of events, which, optionally, also carries data. The propagation of events follows the basic concept of Publish and Subscribe Systems, in [7]. Events are published and middleware ensures that only processes that were associated (or, that were subscribed) for this event will receive them. In [6], the combinations and methods for treating these events correspond to an optimized real-time CORBA model. The model is optimized in its communication interface, aiming to ease the configuration and to manage aperiodic and periodic events. These are a result from underlying layers and they lead to a higher performance due to the robustness of CORBA architecture.

The diversity of applications associated with a CPS, even in a domestic environment, confirms the existence of several media, such as WLAN, ZigBee, LonWorks, and ECHONET [3][18], which difficult the interoperability protocols on different networks. To solve this problem, in [3], a service architecture for control methods is proposed. The proposal asserts that users will be able to control appliances in physical environment through intuitive operations, which react to devices changes in the physical environment. The virtual context approves the changes and performs an action tracking and service execution to a dynamic environment.

 ISBN: 978-1-61208-296-7

In services modeling, in the middleware structure, the unnecessary use of a component is a common occurrence when there is a weak specification of requirements on application execution flows. In [4], the proposal summary of a middleware framework made for monitoring and controlling water distribution systems in large scale is presented. The proposal highlights a middleware framework represented in a layered structure following the FOSD (Feature Oriented Software Design) model. The approach categorizes the components intrinsically or extrinsically, so the middleware can avoid unnecessary use of components in the execution flow.

Our paper describes a middleware modeling, focusing on the sensoring layer. Some of the assumptions aforementioned are discussed first and, then, some CPS characteristics and requirements that are necessary to standardize some processes are defined. This paper also describes the modeling of events abstracted by the middleware, which is capable to adapt itself to the limitations and requirements of a sensor network. The first approach is: i) this paper presents the requirements for these types of systems, ii) it discusses the layered middleware definition and iii) it discusses their implementation by reusing Designs Patterns.

This article is organized as following: In Section II, the description of specifications for cps is presented. In Section III, modeling to event's for middleware. In Section IV, overview of middleware architecture is presented. In Section V, the functional and non-functional requirements for the CPS are discussed. In Section VI, design pattern for middleware is presented. In Section VII, the structure and standardization of data are discussed and, finally, in Section VIII, conclusions and discussions are presented.

II. DESCRIPTION OF SPECIFICATIONS FOR CPS

The description of specifications for modeling middleware begins by determining the heterogeneous set of features, which should be considered when building applications for a CPS.

- *Heterogeneity*: The diversity of sensor elements and actuators supported by a CPS demands a justified attention for the set of functional and non-functional requirements, which should comply with the wide scale of embedded hardware devices that composes the CPS application and actuation environment.
- *Hardware:* The set of devices which comes to be part of the CPS is very extensive, consisting in sensors, mobile phones, tablets, etc. However, the modeling described in this paper restricts itself to the architectures heterogeneity that composes the sensor layer, leaving control activities and actuation in the upper layers, on interface level.
- *Topology*: The topologies supported by the CPS can be various; however, the need for comprisement demands the use of topologies with high availability and reliability characteristics. The IEEE802.11s standard, in [14][15], describes variations of a mesh topology with high level of reliability, especially on the new IEEE802.16, in [16][17]. The mesh networks arise as an alternative to the improvement in wireless communication services. The network control is distributed and the network nodes can transmit and receive data while simultaneously routing through the surrounding nodes. This article restricts itself to the communication scenarios that use mesh and star networks.
- *Communication Protocols:* The diversity of devices appeals to the use of various communication protocols, which maximize the special characteristics of those hardwares. While working with wireless sensors, the set of requirements that a communication protocol should achieve begins with the mobility and low consumption of power. The middleware modeling takes in consideration the multihop and singlehop set of protocols, which predominates by its efficiency on diversity of applications in a wsn (wireless sensor network), and, as a consequence, attends to the dynamicity and comprisement characteristics of the CPS.
- *Communication standards:* The fast grow of applications for wireless communication environments has been motivating the development of various communication standards, which present their own characteristics associated to communication standards of a specific set of devices. The proposed architecture abstracts specific parameters of these communication standards, such as ZigBee, IEEE802.15.4, and others, while offer support for components association through two topologies (mesh and star).
- *Spatiotemporal Variables:* The CPS associate a large set of sensors with heterogeneous characteristics, which demand the specification of temporal requirements. These requirements must provide reliability to the application in execution and also provide the permanent monitoring of sensors, opportune processing of the information received, and the monitoring of data storage spots and events, which trigger processes in other applications.

III. MODELING TO EVENT'S FOR MIDDLEWARE

Inside the middleware, the operations executed through components of the underlying layers are represented as events. For example, when a new *data sink*, which is associated to a specific architecture of hardware, is started along with the middleware, an event of connection is executed and managed by a "capture" method. This method operates as a set of functions that, in this case, would be the connection interface for new devices and also would identify the frame structure and invoke the service that will perform the decoding. The information is stored in a structured manner and it is accessed by the layer of services.

 ISBN: 978-1-61208-296-7

Each event associates a set of spatiotemporal non-functional requirements. Following this statement, we conclude that an event is defined as the combination of one or more event conditions, which are restrictions in terms of attributes, time, and localization, following the model described in [13].

$$PE_{id}\{t_{Eid}, l_{Eid}, V_{Eid}\} \quad (1)$$

where PE_{id} is the physical event identifier, t_{Eid} and l_{Eid} are the spatiotemporal occurrences, and V_{Eid} is the physical event attribute. The observed phenomenon is represented as:

$$PO(MTid, SRid, i)\{t_O^o, l_O^o, v_O\} \quad (2)$$

where PO is the observed physical phenomenon; $SRid$ is the kind of sensor installed on $MTid$ (mote id); i is the observed phenomenon; and $t_{O(MTid,SRid,i)}^{o}$, $l_{O(MTid,SRid,i)}^{o}$, and $v_{O(MTid,SRid,i)}$ are the occurrences of physical observation, time, place, and attributes, respectively.
The sensor event, in [13], serves as the first level for observations in the CPS model of events:

$$SE(MTid, Sid, i)\{t_{SE}^{g}, l_{SE}^{g}, t_{SE}^{eo}, l_{SE}^{eo}, V_{SE}, ps\} \quad (3)$$

$MTid$ is the mote sensor that generates the event based on the identification of Sid sensor event at the instance of i event. Furthermore, a 6th property is used, where $t_{SE(MTid,SRid,i)}^{g}$ and $l_{SE(MTid,SRid,i)}^{g}$ are the time and place related to the occurrence of events and mote sensor, respectively; $t_{SE(MTid,SRid,i)}^{eo}$, $l_{SE(MTid,SRid,i)}^{eo}$, and $V_{SE(MTid,SRid,i)}$ are the estimated occurrence time to event, place and attribute, according to the mote sensor. $ps_{(MTid,SRid,i)}$ is the mote sensor degree of reliability in relation to the event sensor.

If we, for example, simplify the particular representation of an event with the 2.1 formula, the heat monitoring of the x1 server denoted as $SCServer_{x1}$ on $t_{SCServer\_x1}$ instant, located at $l_{SCServer\_x1}$ position, detected by the $V_{SCServer\_Temp\_x1}$ Temperature and $V_{SCServer\_Umi\_x1}$ Relative Humidity attributes can be expressed as following:

$$SCServer_{x1}\{t_{SCServer\_x1}, l_{SCServer\_x1}, [V_{SCServer\_Tem\_x1}, V_{SCServer\_Umi\_x1}]\} \quad (4)$$

Thus, the heat event on server x1 could be detected by the evaluation of Temperature and Relative Humidity attributes of the near sensor (x1), and could also be associated with attributes of other events defined for the system, which may be represented following the (2) and (3) formulas. The middleware events representation and forms of abstraction are described in section IV.

IV. OVERVIEW OF MIDDLEWARE ARCHITECTURE

As represented in Figure 1, the middleware project is divided in layers and the communication happens from the superior layers to the underlying ones. In the process, each layer is responsible for offering a set of needed abstraction and integration functions between the underlying layers, taking in consideration specific requirements of the devices that are part of the application.

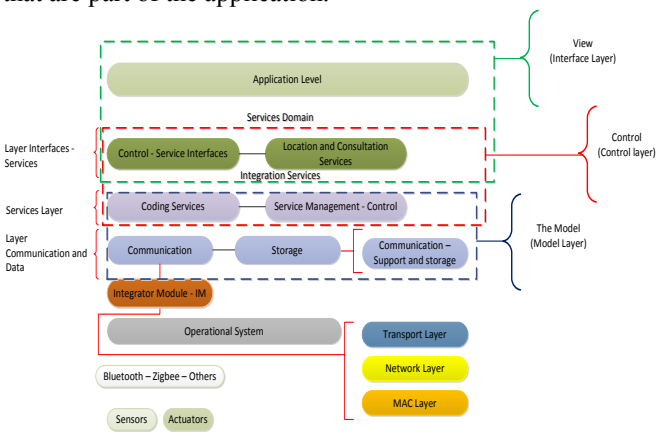


Figure 1. Middleware - Structure Layered

The basic idea for implementation of components according to a layered architecture follows a simple logic: the components defined as modular units with required and provided interfaces are organized in layers, and the components of superior layers communicate with the ones of underlying layers, but the reverse is not true. This method of implementation and communication of components optimizes the proposed middleware logic and it is oriented in the superior layers. In these layers, service requirements follow a hierarchy (the user demands a service to the inferior layers) and the results follow a vertical flow, from bottom to top.

a) Communication and Data Layer (Communications and Support): On this layer, the access to information must be ensured, depending on which role the user is performing during the requisition. Even so, some parameters, which are mentioned in [8], are taken in consideration in the elaboration of this module: dynamic topology, heterogeneous platforms, mixed traffic, power consumption management, and communication protocols. To achieve the quality of these parameters, it will be necessary to standardize the use of topologies and protocols for applications that should be supported by middleware, considering the diversity of CPS backgrounds.

- Integrator module (IM): The most important element for our middleware is called Integrator Module (IM). This IM acts as an element of integration between sensoring devices, actuators (*Mote Sensor T*ransmitter-MST and *Mote Sensor Actuator - MSA*), and middleware. Basically, the IM performs the capture control of network traffic, the insertion and removal of hardware modules, the abstraction of routing protocols, and the interfaces control. The topology characteristics are attended by middleware intuitively by IM, since the organization and routing parameters are transported on the protocol encapsulation at the network and link layers (Figure 1).

 ISBN: 978-1-61208-296-7

To ease the IM access and configuration, it is necessary to develop an API that allows the configuration of new MST values. These values are: total package size, routing package size, and data payload package size.

b) *Service Layer (Management Services):* The construction of applications, divided in communication and application modules, which resides in various computing environments, has been a great step to distributed computation. On the service integration layer, the flexibility is determined in integration with sensors network inside of a CPS, which need to receive standard technologies that offer resources to the development of new applications in return. This paper appeals to the junction and reuse of existing designs pattern and also to the use of web technologies to manage services and data both locally and remotely, considering the following features:

- *Scenario analysis:* this characteristic is indispensable because it allows, along with a requirement analysis, in [9], the determination of language, tools, and techniques that are most adequate to the abstraction of the underlying layers processes, which may satisfy objectively the application requirements to CPS, in [5].
- *Requirements classification:* On the initial requirement analysis, in [10], the total set is classified and considered obligatory, applicable, and not applicable. Following this little structured classification method, the relevance of each requirement is attributed to the computational components integration with applications through a structure based in objects and services for functional, non-functional, and spatiotemporal. This classifying method allows the determination to be made with a higher level of flexibility to relevance, coupling level and abstraction to each requirement.
- *Technologic complexity:* the selection of the most adequate integration technology is very complex due to the diversity of interfaces and programming resources that can be used, in [10]. An inadequate choice may produce serious effects in the way that services will behave at the end.
- *Standardizing*: The user receives access to the information in a consistent and easy to understand manner. Here, the use of MVC (Model View Controller) design pattern, in [11] and [12], and Facade design pattern is proposed. The reuse of such design patterns results from the need for more structured models, which should be robust and, at the same time, flexible to answer to the high CPS scalability and complexity.
- *Process Definition*: Structure definition of each stage in the process with intuit of obtaining a better classification of environment requirements for the integration middleware.

c) *Interface Layer and Domain Service Layer:* On this layer, all software parameters that automate or support the processes performed by the users on the application layer are found and grouped. Those parameters or aspects typically include each task that is part of the processes, as well as the rules and restrictions associated.

The services infrastructure is totally hybrid and should follow the basic model described in three levels:

- *Network*: The middleware architecture considers the wired means of communication, through which it gets connected with the server where the data and management modules are reposing. The wireless mean is used by devices in the acting and sensor layer.
- *Hardware*: on this infrastructure level, the heterogeneity of components and applications that make part of the middleware layer is described. Data base servers, Web and Web service, and SMS.
- *Software*: On software level, it is described the set of applications used in service domain for abstraction of process execution, resources management, and information treatment and storage.

V. FUNCTIONAL AND NON-FUNCTIONAL REQUIREMENTS FOR CPS

a) Functional Requirements: Functional requirements (Figure 2) define, in a concrete method, what the middleware should do, and they will depend on the characteristics of the application to which they are being developed, a set of minimal *requirements are presented in F*igure 2.

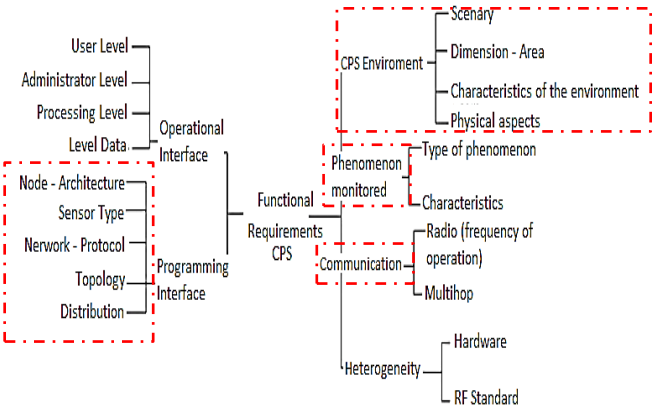


Figure 2: Functional Requirements for middleware modeling

b) Non-Functional Requirements: Before determining the set of non-functional requirements that make part of an application for CPS, it is relevant to identify the set of elements that make part of the operational interface of the hardware architecture associated to the application. As our case is a wireless sensors network, the recognition of this set of parameters allows the identification of non-functional requirements, which result of the hardware and application association, in a more clear way (Figure 3).

c) Non-Functional Temporal Requirements: The CPS associates a large set of sensors with heterogeneous characteristics. These characteristics demand a specification of temporal requirements that are able to provide reliability

 ISBN: 978-1-61208-296-7

to the application in execution and that provide the opportune processing of information received and the monitoring of permanent sensors, storage spots of information, and events that trigger processes and other applications.

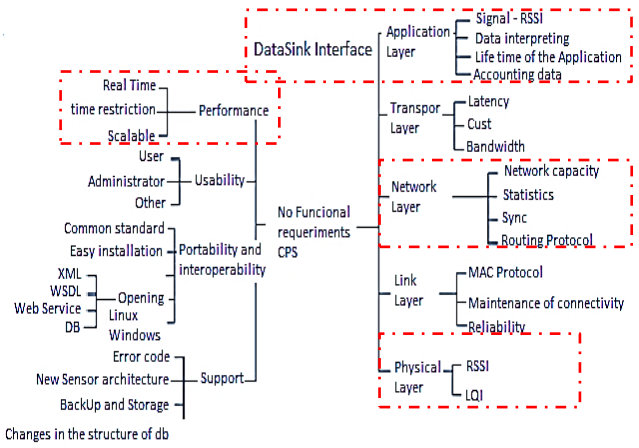


Figure 3: Non-Functional Requirements for middleware modeling

The IM connects itself to the central server and starts a connection thread with the middleware. When the connection to the server interface (data layer) is established, a second thread starts. The second thread acts as a server thread that invokes one or more worker threads.

Worker threads manage the interfaces for each data sink connected to the IM (in practice, the IM has both functions). When the first two threads start, a third one starts then and allows control and traffic classification, classifying the package flow by the communication protocol (*multihop and singlehop*) used by the sensor node (MT), as well as the type of information that it carries in the data payload, applying (6) and (7) In the classification, some non-functional requirements, non-functional spatiotemporal, and attributes extracted from the data structure entered by the data sink for IM are considered, applying (5)

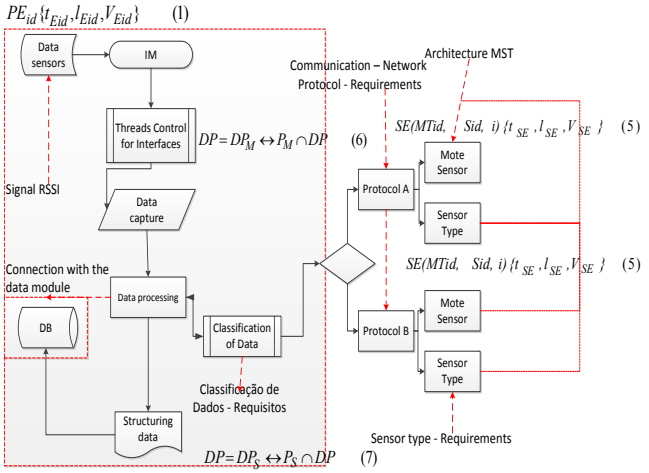


Figure 4: Communication Data Sink - Storage - Data Management Service - Events and Requirements

Data structure information of each data is presented as events that include non-functional spatiotemporal requirements to attribute reliability to the system. The set of events is managed by the service layer. Those sensor events and attributes are represented through a (3) simplification.

$$SE(MTid,Sid,i)\{t_{SE}, l_{SE}, V_{SE}\} \quad (5)$$

Where SE is the sensor event, MTid corresponds to the mote architecture, Sid is the kind of sensor over the mote architecture, i is the event observed by the architecture, and $\{t_{SE}, l_{SE}, V_{SE}\}$ corresponds to the time, space, and attributes requirements, respectively.

The data classification process is performed determining common variables in the multihop protocols structure; this set of variables is represented as P_M, for that case, the DP input data package is classified as a DP_M multihop package when its structure includes a P_M set of values. The same logic applies to DP_S singlehop.

$$DP = DP_M \leftrightarrow P_M \cap DP \quad (6)$$

$$DP = DP_S \leftrightarrow P_S \cap DP \quad (7)$$

The CPS presents a high complexity and heterogeneity of hardware components and applications, which should interact with the physical environment observing, processing, and reacting to external phenomenon in real time. Following those requisitions, it is necessary to make an association with the events model (described in section II) and temporal requirements of periodic character: *i) according to the activation scheme* and *ii) according to execution time*.

- *According to the activation scheme*: These requirements could be of periodic or non-periodic activation. The periodic activation requirements are set one per T period, and exactly on each T period (Figure 5).

Inside the CPS, the wsn needs to use sampling techniques to offer a greater efficiency and, mainly, reliability of the data collected during the monitoring. On the middleware, among the temporal requirements, we have a differentiated and fixed number of samples collected per sensor. For example, considering the use of two routing protocols, singlehop and multihop, to equilibrate the efficiency and reliability of the acquired data in the physical environment, the sensors transmit the result of 5 to 10 readings in a single frame on each T. In practice, we would have 5 to 10 readings at each 100ms with a transmission cycle of up to 1000ms. This last parameter can be altered by the administrator according to the characteristics of the application. The possibility of altering the transmission cycle is interesting for some applications; when there is already a great set of samples per frame, there is no need for a high periodicity transmission.

 ISBN: 978-1-61208-296-7

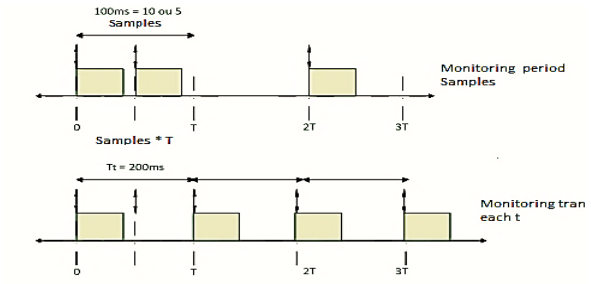


Figure 5: Representation of requirements for periodic activation.

As mentioned before, some applications in CPS, such as the control events, in [7] and [13], may demand the use of nodes to perform observations and periodic transmissions of the detected events characteristics. The degree of correlation between the consecutive mediations of the sensor node may vary according to characteristics of the physical phenomenon temporal variation, with which the CPS interacts.

In the middleware project for the CPS, it is necessary to flexible the rigor of some requirements execution, leaving the objectives of each one of them to be reached, also, in function of the best effort. Here, it is needed to consider the application of temporal requirements according to the execution time.

For that, it is possible to use some combinations of requirements, like periodicity, execution and validation time, and even the junction of temporal and functional requirements, which demands a lower level of precision, but allows the conclusion of the process. Inside the diversity of applications in CPS, the delivery and reception of data is what matters in some situations, and not the time spent in the route of reception and transmission.

- *Temporal Requirement according to execution time:* On the application of temporal requirements according to the execution time (Figure 6), it is necessary to emphasize that there are applications such as the ones of critical systems that does not admit, under any circumstances, the violation of the event implementation period.

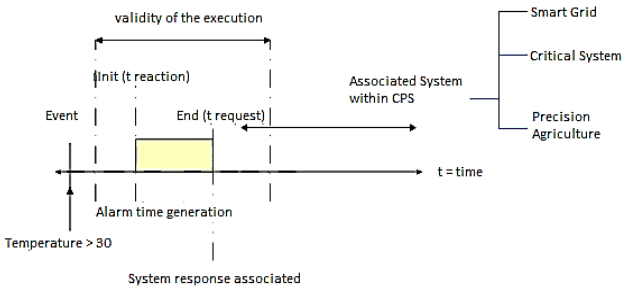


Figure 6: Representation of temporal requirements according to execution time.

The functional and non-functional requirements presented are the result of a careful study of the CPS characteristics and the wsn operating structure. The choice and determination of new requirements must pass through a specific and detailed assessment of the current architectures for sensor networks, and of spatial and physical characteristics of integration and interaction with the physical environment, which are common to all CPS.

VI. DESIGN PATTERN FOR MIDDLEWARE

The implementation of a new or reused existing design pattern to build a middleware for CPS results in an extensive survey of new issues due to the high complexity of the application scenario that presents this new concept of systems. The construction of CPS follows a distributed logic and, as a consequence, the proposed structure for managing the diversity of components and services that exist inside of a CPS cannot be different.

The middleware initial structure must follow design patterns that allow a comprehensive description and functional requirements expression of the CPS customer and the representation of the predictable behavior of such application in execution time. The purpose of using MVC (*Model-View-Controller)* and facade is to explore the advantages of existing *design patterns* to produce a reusable middleware structure and easy maintenance for a CPS on sensor level. The MVC model adopted is organized in three layers (Figure 7), in which the intermediate layer exerts the control services and determines the application behavior.

Figure 7 shows the three-layer MVC model, which integrates the Facade Design Pattern that attributes important features to the middleware for reuse and easy maintenance. Both are models of object-oriented design that give a greater level of coupling and allow the modular modeling of each middleware components, which are associated through a service layer. Figure 7 presents the macro modeling architecture that describes the Model, the View, and the service, and also includes the DAO (Data Access Object) design pattern used to determine the access rules to data structured system.

The *design patterns* used are an excellent option to manage the middleware components in lower layers, providing a modeling that is structured and flexible enough to be reused by services architecture at the highest middleware level.

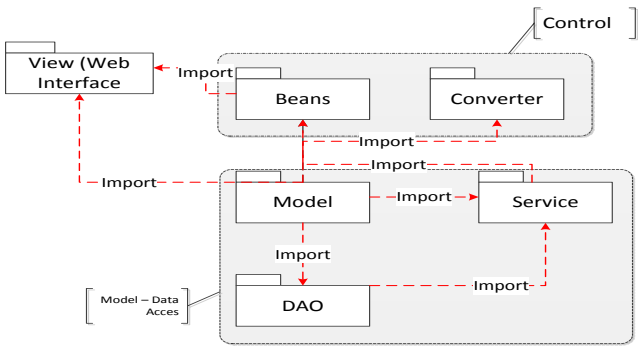


Figure 7: Specification diagram - packages abstraction

 ISBN: 978-1-61208-296-7

The dependency between components is defined as a result of the model that was used. The object-oriented Designs Patterns require a higher level of coupling between components, what is ideal for components that perform the abstraction and communication of data on the underlying layers, and arise from the diversity of architectures and protocols that are part of CPS. The strong dependence on middleware lower layers favors the system robustness, which is required by the high granularity of data resulted from these layers.

The coupling on upper layers is associated with service-oriented Design Pattern to web technologies, allowing the determination of some services engagement levels that are best suited for certain components, specifically those necessary to promote the reliability and availability of the system, positioning the service as a resource highly available and accessible, inclusively being required by the high willingness to critical events in the CPS.

VII. STRUCTURE AND STANDARDIZATION OF DATA

The information is defined as structured in a relational database, however, given its heterogeneity, it is impossible to put all the information into a carefully specified database. Routing information and others that are specific to hardware are collected casually, before it has become known how they will be managed. This type of data should provide, in some cases, a variable structure. Some of the attributes contained in such data can be shared between existing entities, but some attributes may exist only in some entities. These are classified as semistructured data. The structured model and semistructured data suggest the file structure adjustment to a format that can be shared and reused by other applications. This is the case of XML data language. XML files structures are suggested for middleware, extracting information from structured database that determines common attributes of the used communication protocols and collected information inside of the CPS. When there is some set of unstructured data, these can be stored on a standard structure, so they can be analyzed later.

VIII. CONCLUSION AND DISCUSSIONS

Although there are increasing researches on CPS, the results achieved in relation to this new concept of hybrid systems are few. The reason for this is that we still cannot count on a mature science to support the features and requirements of CPS engineering systems with a high level of reliance, and the consequences are profound. The current analysis tools are unable to cope with the high level of CPS complexity, or at least adequately predict the behavior of these systems. A clear example would be the recent events in the electrical grid, blackouts on large regions as a result of widespread failures in the system. This probably illustrates the shortcomings and limitations of current technologies, which complicates the conceptualization and creation of independent projects of engineering systems that interact with the natural and the physical world. At the same time, the demands of today's governments to develop renewable technologies, such as energy, advanced manufacturing of smart materials, vehicles powered by electricity, air transportation etc., create unprecedented opportunities to rethink many of existing systems as parts of a CPS.

The proposal of work introduced in [1] presents some lacks in relation to the integration of communications standards specified in the physical layer to support the heterogeneity of communication patterns and determine the scaling characteristic of devices in WebMed. Unlike that, the middleware architecture proposed and described in this paper intends to concentrate the devices heterogeneity result, which exists on CPS, through the middleware's communication layer that couples the IM module. And it also intends to concentrate the management of scalability aspects that result from the adhesion of heterogeneous hardware in current implementation. For this porpoise, this paper has started describing the layers of the proposed middleware and the set of services restricted to each layer using design patterns to implementation, which is called MVC. Creating middleware structures using existing design patterns favors easy and quick system maintenance and can be extended beyond sensoring level. The middleware implementation using MVC, facade, and DAO design patterns facilitates the implementation and alteration of each layer's functionalities, allowing the middleware to behave itself as a supervisory system on the application layer.

On CPS, the construction of each event and corresponding action happens and uses resources in a particular context. This information is fundamental to know which agents should be included in the action flow, so that the desired task is corrected in its own context, considering restrictions offered by the CPS application environment. The formalization and modeling of events, adapted from [13] and represented along with requirements according to the activation scheme and execution time, are easily abstracted from the user by design patterns without affecting the performance and transparency of the application.

The Design Patterns proposed in this paper for the middleware development follow the object-oriented and services models to balance the coupling of components and save on functional and non-functional requirements to reduce processing demands and application energy. The CORBA architectures are characterized by the variety of applications that can be created and integrated on them. However, their performance requirements, their excessive exchange of messages between components to complete a process, or their interruption on execution time let the architecture restricted to application scenarios with low mobility and greater processing resource. Therefore, CORBA architecture is not a good alternative to applications that manage hardware architectures with restricted characteristics.

 ISBN: 978-1-61208-296-7

REFERENCES

[1] Hoang, d. D.; Paik, H.-Y.; Kim, C.-K. “Service-oriented middleware architectures for cyber-physical systems”. IJCSNS International Journal of Computer Science and Network Security, vol. 12, no. 1, pp. 79–87, Jan 2012.

[2] Park, S. O.; Do, T. H.; Jeong, Y.-S.; Kim, S. J. “A dynamic control middleware for cyber physical systems on an ipv6-based global network”. International Journal of Communication Systems, John Wiley and Sons, Ltd, pp. n/a–n/a, Dec 2011. doi: 10.1002/dac.1382.

[3] Lai, C.-F.; Ma, Y.-W.; Chang, S.-Y.; Chao, H.-C.; Huang, Y.-M. “Osgi-based services architecture for cyber-physical home control systems”. Computer Communications, vol. 34, n. 2, pp. 184 – 191, February 2011. http://dx.doi.org/10.1016/j.comcom.2010.03.034.

[4] Mudasser, I.; Beng, L. H. “A cyber-physical middleware framework for continuous monitoring of water distribution systems”. In: Proceedings of the 7th ACM Conference on Embedded Networked Sensor Systems (SenSys ’09). New York, USA:, pp. 401–402. 2009. doi=10.1145/1644038.1644122

[5] Wood, A.; Stankovic, J.; Virone, G.; Selavo, L.; Zhimin H.; Qiuhua C; Thao D; Yafeng W; Lei F; Stoleru, R.; , "Context-aware wireless sensor networks for assisted living and residential monitoring," Network, IEEE , vol.22, no.4, pp.26-33, July-Aug. 2008. doi: 10.1109/MNET.2008.4579768. URL: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=4579768&isnumber=4579761

[6] Zhang, Y.; Gill, C.; Lu, C. “Reconfigurable real-time middleware for distributed cyberphysical systems with aperiodic events”. Distributed Computing Systems, International Conference on, IEEE Computer Society, Los Alamitos, CA, USA, vol. 0, pp. 581–588, 2008. doi:10.1109/ICDCS.2008.96

[7] Eugster, P. T.; Felber, P. A.; Guerraoui, and R.; Kermarrec, A.-M. “The many faces of publish/subscribe”. ACM Comput. Surv., New York, USA, vol. 35, no. 2, pp. 114–131, June. 2003. doi.acm.org/10.1145/857076.857078.

[8] Xia, F.; Ma, L.; Dong, J.; Sun, Y. “Network QoS management in cyber-physical systems”. In: Proceedings International Conference on Embedded Software and Systems Symposia. Washington, DC, USA: IEEE Computer Society (ICESSSYMPOSIA ’08). pp. 302–307. July 2008. 10.1109/ICESS.Symposia.2008.84.

[9] Romer, K.; Mattern, F.; , "The design space of wireless sensor networks," Wireless Communications, IEEE , vol.11, no.6, pp. 54- 61, Dec. 2004. doi: 10.1109/MWC.2004.1368897.

[10] Cheng, B.H.C.; Atlee, J.M.; , "Research Directions in Requirements Engineering," Future of Software Engineering, 2007. FOSE '07 , vol., no., pp.285-303, 23-25 May 2007.doi: 10.1109/FOSE.2007.17.

[11] Mcheick, H.; Yan Qi; "Dependency of components in MVC distributed architecture”. Electrical and Computer Engineering (CCECE), 2011 24th Canadian Conference on , vol., no., pp.000691-000694, 8-11 May 2011 doi: 10.1109/CCECE.2011.6030542.

[12] Tao P; Lianying S; Hong B; , "Design and implementation of ATM simulation system based on MVC pattern". Educational and Information Technology (ICEIT), 2010 International Conference on, vol.1, no., pp.V1-328-V1-331, 17-19 Sept. 2010. doi: 10.1109/ICEIT.2010.5607693.

[13] Tan, Y; Vuran, M. C. and Goddard, S, “Spatio-Temporal Event Model for Cyber-Physical Systems”, In Proceedings of the 2009 29th IEEE International Conference on Distributed Computing Systems Workshops ICDCSW (2009). pp 44-50. Nov 2009. doi:10.1109/ICDCSW.2009.82

[14] Shyy, D.J.; “Military Usage Scenario and IEEE 802.11s Mesh Networking Standard”. Military Communications Conference, MILCOM 2006. IEEE , vol., no., pp.1-7, 23-25 Oct. 2006. doi: 10.1109/MILCOM.2006.302448

[15] Safonov, A.; Lyakhov, A. and Sharov, S.; “Synchronization and beaconing in IEEE 802.11s mesh networks”. Telecommunications, ICT 2008. International Conference, vol., no., pp.1-6, 16-19 June 2008. doi: 10.1109/ICTEL.2008.4652690

[16] Tae-Hwan K and In-Cheol P; “Low-Power and High-Accurate Synchronization for IEEE 802.16d Systems”. Very Large Scale Integration (VLSI) Systems, IEEE Transactions on, vol.16, no.12, pp.1620-1630, Dec. 2008. doi: 10.1109/TVLSI.2008.2001567

[17] Min-Seok Kang; Jaeshin Jang; “Performance evaluation of IEEE 802.16d ARQ algorithms with NS-2 simulator”. Communications, APCC '06. Asia-Pacific Conference on , vol., no., pp.1-5, Aug. 2006 doi: 10.1109/APCC.2006.255785

[18] Garay, Jorge R. B and Kofuji, S.T., “Architecture for sensor networks in cyber-physical system”. Communications (LATINCOM), 2010 IEEE Latin-American Conference on , vol., no., pp.1-6, Sept. 2010 doi:10.1109/LATINCOM.2010.5641126.

Jorge R. Beingolea Garay is B.Sc. in Computer Science (2004) - Universidad Inca Garcilaso de La Vega, M.Sc. in Electrical Engineering (2007) - University of São Paulo, and PhD. in Electrical Engineering (2012) - University of São Paulo. He is currently a Group PAD Researcher in the Integrated Systems Laboratory (LSI) at EPUSP. He has experience in Electrical Engineering, with emphasis on Wireless Sensor Networks, Wireless Communication, Computer Architectures, Wireless Networks and Pervasive Computing, Service-Based Architectures, and Cyber-Physical System.

Sérgio T. Kofuji is B.Sc. in Physics - University of São Paulo (1985), M.Sc. in Electric Engineering - University of São Paulo (1988), and PhD. in Electric Engineering - University of São Paulo (1995). He is currently a Group Chief Researcher in PAD of the Integrated Systems Laboratory (LSI) at EPUSP.

Alexandre M. De Oliveira is B.Sc. in Electrical Engineering with Computer emphasis - Catholic University of Santos (2008), M.Sc. in Electrical Engineering (2012) - University of São Paulo, and a graduate student in the University of São Paulo. He is currently Group PAD Researcher in the Integrated Systems Laboratory (LSI) at EPUSP and professor at Unimonte and Faculdade Praia Grande, with experience in Electrical Engineering, with emphasis on VLSI Design, UWB I-Radar, Timed-array propagation.

Lensless Ultra-Miniature CMOS Computational Imagers and Sensors

David G. Stork and Patrick R. Gill
Computational Sensing and Imaging
Rambus Labs

***Abstract*—We describe a new class of lensless, ultra-miniature computational imagers and image sensors employing special optical phase gratings integrated with CMOS photodetector matrices. Because such imagers have no lens, they are ultra-miniature ($\sim$100 μm), have large effective depth of field (1 mm to infinity), and are very inexpensive (a few Euro cents). The grating acts as a two-dimensional visual "chirp" and preserves image power throughout the Fourier plane (and hence preserves image information); the final digital image is not captured as in a traditional camera but instead computed from raw photodetector signals. The novel representation at the photodetectors demands that algorithms such as deconvolution, Bayesian estimation, or matrix inversion with Tikhonov regularization be used to compute the image, each having different bandwidth, space and computational complexities for a given image fidelity. Such imaging architectures can also be tailored to extract application-specific information or compute decisions (rather than compute an image) based on the optical signal. In most cases, both the phase grating and the signal processing can be optimized for the information in the visual field and the task at hand. Our sensor design methodology relies on modular parallel and computationally efficient software tools for simulating optical diffraction, for CAD design and layout of gratings themselves, and for sensor signal processing. These sensors are so small they should find use in endoscopy, medical sensing, machine inspection, surveillance and the Internet of Things, and are so inexpensive that they should find use in distributed network applications and in a number of single-use scenarios, for instance in military theaters and hazardous natural and industrial conditions.**

Keywords: Computational sensing, phase grating, diffractive imager, application-specific sensing

I. Introduction

The traditional *camera obscura* model—in which each point in the scene is imaged onto a single point on a sensor or image plane—has dominated the science and technology of imaging systems for several millennia, at least for sources illuminated by incoherent light. The Chinese philosopher Mo Ti traced an inverted image produced by a pinhole camera to record an image in the fifth century B.C. [1] and Johannes Kepler traced a real image projected by a converging lens onto paper in 1603. Chemical recording of projected images, such as by mercury or silver halide, was invented in 1826 and the first true digital camera was built in 1975, [2] all these exploiting the fundamental camera obscura architecture.

The rise in digital imaging, where image processing can be incorporated into the data chain, has enabled new imaging architectures. Cathey and Dowski took an early and conceptually important step away from the traditional camera obscura model by exploiting digital processing [3]. They designed a cubic-phase optical plate which, when inserted into the optical path of a traditional camera, led to an image whose (significant) blur was independent of the object depth: the image on the sensor plane did not "look good" as it would in a traditional camera obscura. Subsequent image processing sharpened the entire blurred image, thus leading to enhanced depth of field. Since then the field of *computational imaging* has explored imaging architectures in which the raw signals do not superficially resemble a traditional image; instead, the final image is computed from such signals. In this way, many optical aberrations can be corrected computationally rather than optically. More and more of the total imaging "burden" is borne by computation, thereby expanding the class of usable optical components. This imaging paradigm has led to new conceptual foundations of joint design of optics and image processing, [4] as well as a wide range of non-standard imaging architectures such as plenoptic, coded-aperture and multi-aperture systems, each with associated methods of signal processing [5]–[9].

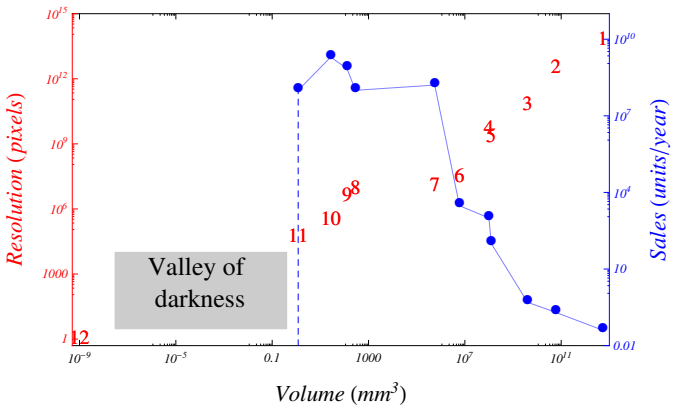


Fig. 1. The resolution (in pixels) versus the physical volume (in mm^3) of representative lens- and mirror-based telescopes and cameras (log-log scale). Notice there is a seven-order-of-magnitude range in physical volume devoid of such cameras (the **Valley of darkness**). 1 Grand Canaria telescope, 2 Hubble telescope, 3 1-m telescope, 4 30-cm telescope, 5 AWARE 2 camera, 6 Professional camera,7 Consumer DSLR, 8 iPhone 5 camera, 9 Pelican camera, 10 Miniature VGA, 11 Medigus camera, 12 Single photodiode (without lens). The blue points indicate the sales of representative imagers of different physical volumes in units/year worldwide in 2013. (The unit sales figures are estimates based on historical data and market reports and do not include research prototypes and unreleased products.) Note that there is a precipitous drop in sales at the Valley of darkness. Our lensless integrated diffraction grating/CMOS imagers lie within this "valley."

The economic pressures for miniaturization of electronic devices, including cameras, arising in the mobile computing market have led to smaller imager form factors [10]. Figure 1 shows the resolution, in total pixels per exposure, versus physical volume of imaging systems in the traditional camera obscura architecture (or curved mirror equivalent). While such

 ISBN: 978-1-61208-296-7

imagers span 22 orders of magnitude in physical volume and 15 orders of magnitude in pixel resolution, the smaller the imager the greater the number sold commercially... but only down to a scale of roughly 1 mm^3. There is a conspicuous gap of seven orders of magnitude in physical volume—the "Valley of darkness"—between the smallest digital camera and a single unlensed photoreceptor. It seems that the camera obscura model has reached its physical limits and cannot be scaled much smaller. A new imaging architecture is required to span the Valley of darkness.

Recently, a new miniature imaging architecture has been explored, one based on integrating optics with CMOS photodetectors [11]–[14]. In brief, this architecture forgoes lenses and relies instead on simple square-wave diffraction gratings created in CMOS itself. The earliest designs in this architecture relied on CMOS wires to act as amplitude optical grating patches, the gratings producing a wavelet-like representation of the scene on the sensor matrix. More recently, square-wave *phase* gratings have also been explored [15]. For a given image resolution, such diffractive elements enable the construction of imagers much smaller than does the basic camera obscura model. (We mention in passing that related CMOS structures have been explored for integrated spectroscopy as well [16].)

There are a number of limitations of such previous work. First, amplitude gratings based on CMOS wires have poor low-light sensitivity because most of the incident light never strikes the photodetector. Second, regular diffraction gratings are by their very nature wavelength sensitive, i.e., the pattern of light on the photodetectors depends strongly upon the wavelength of incident light. Third, such imagers are sensitive to manufacturing defects—specifically a small deviation in the thickness of the grating layer can lead to a large (and difficult to correct) alteration of the diffraction pattern on the photodetectors [13].

The method we describe here, while based on integrated silicate phase optics and CMOS image sensors, is fundamentally different from prior work in a number of deep ways. Our method relies on novel special phase anti-symmetric spiral phase gratings, which overcome prior limitations and afford new functionality [17]. Moreover, our new sensor architecture enables the construction of new classes of ultra-miniature sensors whose output is an estimation of some property of the scene (e.g., visual motion) or a decision (e.g., face detection or barcode reading).

We begin in Section II with a discussion of our fundamental technology and turn in Section III to a short description of our software design and analysis tools. We mention a few application areas for such sensors and imagers in Section IV and conclude in Section V with a brief summary and suggestions for future research.

II. Sensor Optics and Technology

The following description of our sensor technology follows the data path—from target source through diffractive optics to photodetector to digital signal processing.

A. Optics of one-dimensional phase anti-symmetric gratings

The fundamental optical elements employed by our sensors are based on a new type of phase grating having phase anti-symmetry. Figure 2 shows a cross section through our silicate binary phase grating, here specified by three free parameters, w_0, w_1 and w_2 [18]. (Generalizations to more free parameters and multiple thicknesses are straightforward.) Consider point **P** lying on the grating's plane of odd symmetry. Light from each position on one side of the plane is cancelled via destructive interference by light from the symmetric position on the other side of the plane because those waves arrive out of phase. Note that such cancellation occurs regardless of the vertical depth of **P**. As such, all points along the red dashed line are dark; we call this plane an "optical curtain" or simply "curtain" [19], [20]. The location of the curtain on the sensor matrix below does not change despite manufacturing errors in overall grating thickness. Finally, as the angle of incidence of the light changes, the curtains tip by the same angle (Fig. 3).

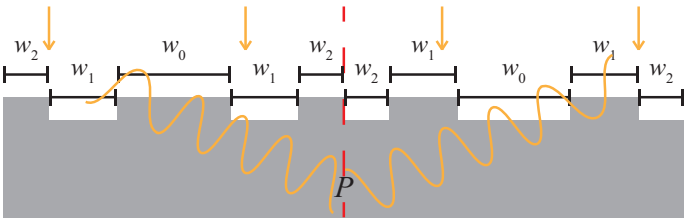


Fig. 2. A cross section through a silicate binary phase anti-symmetric phase grating, where the plane of odd symmetry is marked with a dashed red line. The parameters w_0, w_1 and w_2 describe the surface profile. For the medium's index of refraction n, the step height is chosen to corresponds to optical phase delay of π radians along the red dashed line or "curtain." For such a phase anti-symmetic grating, curtains exist even if the incident light is not normal (Figs. 3 and 4).

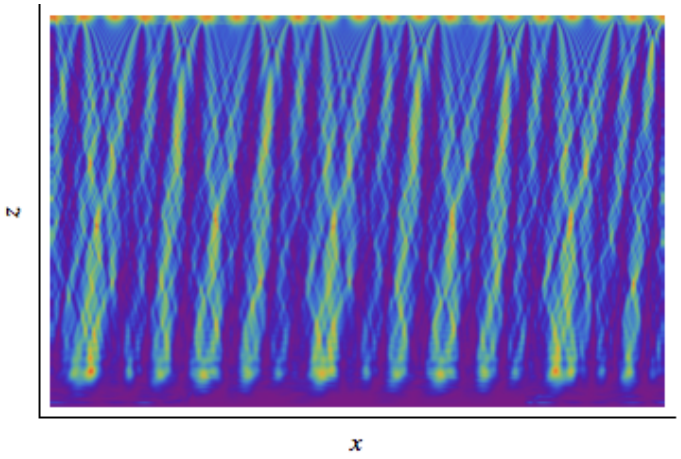


Fig. 3. A finite-difference wave simulation of the electric field energy density for monochromatic light incident at 3.5° passing through an phase anti-symmetric grating where x denotes the position left-to-right and z the depth within the silicate medium. The curtains lie beneath the points of odd symmetry (purple) and are tipped at the same angle as the incident light. Such curtains are invariant to the wavelength of incident light. The photodetector matrix (not shown) lies along the bottom.

B. Phase anti-symmetric spiral gratings

Image sensors are two-dimensional and therefore the phase anti-symmetric grating just described must be generalized to two dimensions. Moreover, two-dimensional gratings must include segments at every orientation so as to sample the

 ISBN: 978-1-61208-296-7

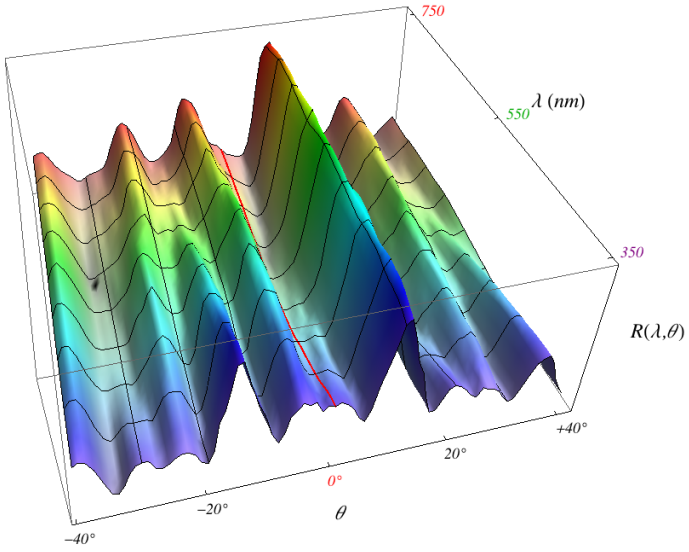


Fig. 4. The response of a single photodetector (pixel) beneath a phase anti-symmetric grating (such as **P** in Fig. 2) as a function of angle of incident light, θ, and wavelength of light, λ. Notice that for normally incident light ($\theta = 0°$) the response nearly vanishes at all wavelengths and that at each incident orientation, the response is nearly invariant with respect to wavelength. The specific form of this response function depends upon the profiles of the grating (described by w_is), which can be tailored to extract information most appropriate to particular applications, including non-imaging applications.

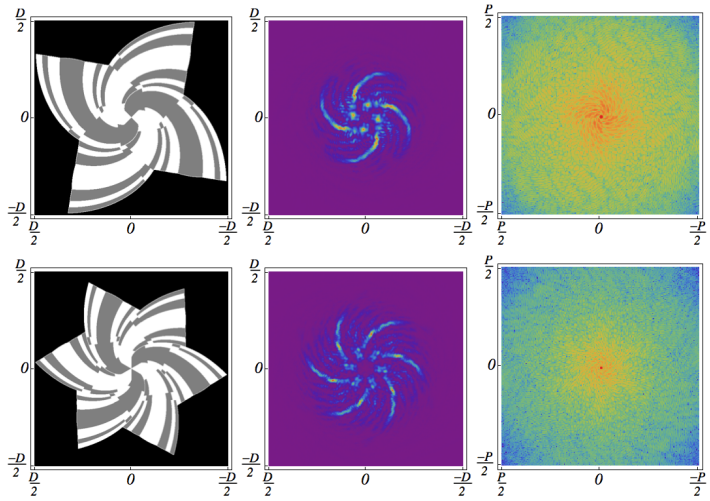


Fig. 5. The left column shows phase anti-symmetric spiral binary gratings, the middle column the point-spread function each produces (both figures of spatial extent $D \times D$, for some distance D). The right column shows the corresponding modulation transfer function (modulus of the Fourier transform) of extent $1/P \times 1/P$, where P is the pixel pitch and determines the Nyquist rate. The top row corresponds to four-fold chiral symmetry and the bottom row corresponds to six-fold chiral symmetry.

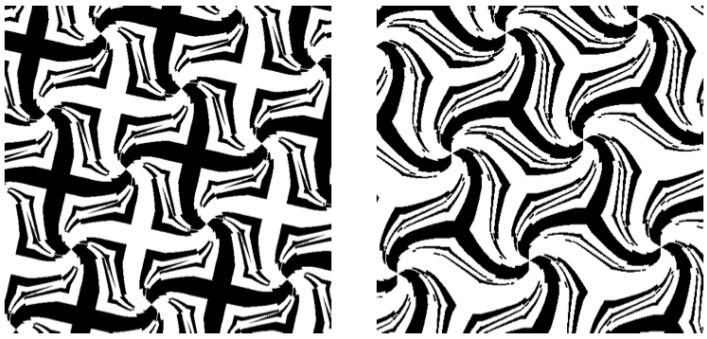


Fig. 6. The individual grating tiles of Fig. 5 can be packed to cover a photodetector matrix of arbitrary area.

Fourier domain uniformly (and possess no zeros) and thereby enable computational reconstruction of the image from sensor responses. Figure 5 shows two examples of basic spiral grating tiles—having four-fold and six-fold chiral symmetry. These spiral grating tiles are constructed by sweeping one-dimensional phase anti-symmetric gratings perpendicularly along the length of each spiral arm. The phase anti-symmetric gratings are lengthened and made more complicated (use more ws) to cover the full tile area and feasible Fourier domain. Both spiral gratings pass information at all orientations and spatial frequencies up to the Nyquist limit, and can be tiled to cover a full photodetector matrix of arbitrary area (Fig. 6) [17]. In actual sensors, incident light covers an area at least as large as that of a full individual tile element.

The wave optics described above assumes the incident illumination is plane-wave. In such a case the pattern of light produced by a grating does not depend upon the distance of the object, so long as the object is farther from the sensor than roughly 10 times the spatial scale of the sensor itself. As such, our sensor has extremely large effective depth of field, from roughly 1 mm to infinity.

The pattern of light produced by the diffraction grating strikes the CMOS photodetector matrix beneath and the signals are sent off chip for digital processing.

C. Signal processing

Sensed signals in our sensor do not resemble an image in a camera obscura but must be processed to yield a digital image. We assume the overall forward imaging model is described by:

$$\mathbf{y} = \mathbf{M}\mathbf{x} + \mathbf{n}, \tag{1}$$

where $\mathbf{y}$ is the vector of photodetector pixel responses, $\mathbf{x}$ is a vector of inputs, $\mathbf{M}$ the matrix describing the linear transformation performed by the two-dimensional optical grating, and $\mathbf{n}$ is additive noise, which describes photodetector noise, Poisson photon statistics, quantization noise, etc. (Other models, such as simple multiplicative noise, could also be assumed.) Then estimation of the input—that is, the reconstruction of the image—is given by:

$$\hat{\mathbf{x}} = \left(\mathbf{M}^t\mathbf{M} + \gamma\mathbf{I}\right)^{-1}\mathbf{M}^t\mathbf{y}, \tag{2}$$

where γ is a Tikhonov regularization parameter and $\mathbf{I}$ the identity matrix [17], [21]. Such estimation is well-conditioned and has higher fidelity when the modulation transfer function of the optical element contains no zeros, as is ensured by our special anti-symmetry phase gratings. Other reconstruction methods include inverse Wiener filtering and Bayesian methods such as Richardson-Lucy deconvolution, [22] each with computational complexities and fidelities that depend upon the accuracy of prior information about the source and other

 ISBN: 978-1-61208-296-7

parameters. Figure 7 shows the estimation of an image through simple matrix inversion with Tikhonov regularization.

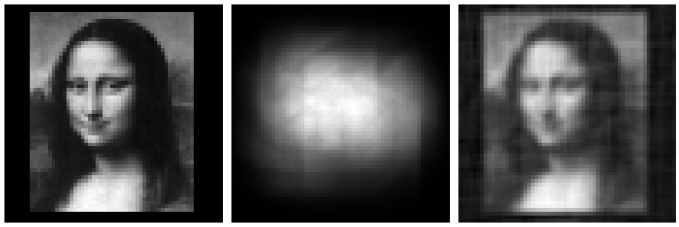


Fig. 7. Image sensing and computational reconstruction of Leonardo's *Mona Lisa* from a lensless phase anti-symmetric spiral phase grating sensor. (Left) The input image. (Middle) The simulated response on the photodetectors due to the six-fold grating in Fig. 5, and (right) the reconstruction by Eq. 2. This image estimate is of higher fidelity than the estimate based on traditional square-wave gratings and photodetector arrays of comparable number of pixels and overall noise level described in earlier work.

III. Simulation/design tools and methodology

Our sensor system design and analysis methods are based on a modular architecture comprising three software tools, all written in *Matlab* and executed on a large network of PCs:

- **Optics of phase gratings**: We simulate the interaction of light with gratings, for instance by finite-difference wave algorithms. These simulations predict the response of physical photodetector pixels to light incident at arbitrary angles.
- **CAD design of gratings and tiles**: We design gratings (spiral and otherwise) and their tilings. The output of our design is either *Matlab*-compatible files for optics simulations or *GDSII* for silicon grating manufacture.
- **Sensor signal processing**: We continue to write our own image reconstruction, signal estimation and pattern recognition software in *Matlab*, often using standard libraries of matrix operations.

We can employ *Perl* software wrappers for these components in order to efficiently design and model the system's end-to-end performance.

IV. Applications

There are many promising applications for our sensors which fall into a number of general categories. A small but important subset of such categories follows.

A. Imaging

The ultra-miniature size of our imagers and sensors make them especially appropriate for medical and industrial endoscopy as well as traditional and novel mobile computing devices. There are many surveillance applications that would profit from low- to mid-level resolutions as well. Because these sensors are so inexpensive (in bulk)—each less expensive than a single frame of 35-mm photographic film—they could find application in a number of one-use imaging scenarios arising in military theaters, hazardous industrial conditions (crash tests) and natural environments. Another general area is inexpensive mobile medical imaging and sensing.

B. Motion estimation

The optical gratings and signal processing algorithms can be tailored to broad image sensing applications such as occupancy detection for controlled lighting, motion (motion-activated devices), visual looming (pre-impact automotive airbag deployment), interactive toys, and numerous applications in support of the Internet of Things [23].

C. Pattern recognition

These sensors can extract informative visual information for pattern recognition applications, such as face detection (authentication), one-dimensional barcode and two-dimensional QR code reading, gesture recognition and many others. Of course, the signal processing is then based on principles of pattern recognition appropriate for the task at hand [24].

V. Conclusion

We have designed and verified through full end-to-end system simulation a new class of lensless computational imagers based on phase anti-symmetric spiral gratings. These imagers promise to be smaller (lower physical volume) than any existing lens-based imagers of comparable resolution, very inexpensive, and customizable to both imaging and a wide range of sensing and image measurement tasks.

Hardware implementations of our computational sensors are underway and practical fielded applications will lead to many interesting problems in efficient application-specific algorithms, either on special-purpose ASICs, on highly parallel graphics processor units (GPUs), or on general-purpose central processor units (CPUs). Networks of such sensors highlight several problems and opportunities in power usage and bandwidth optimization.

Acknowledgment

We thank Thomas Vogelsang and Michael Ching for helpful comments.

References

[1] T. Gustavson, *Camera: A history of photography from Daguerreotype to digital*. New York, NY: Sterling Publishing Co., 2009.

[2] D. Wooters and T. Mulligan, *A history of photography—from 1839 to the present*. New York, NY: Taschen, 2005.

[3] W. T. Cathey and E. R. Dowski, Jr., "A new paradigm for imaging systems," *Applied Optics*, vol. 42, no. 29, pp. 6080–6092, 2002.

[4] D. G. Stork and M. D. Robinson, "Theoretical foundations of joint design of electro-optical imaging systems," *Applied Optics*, vol. 47, no. 10, pp. B64–75, 2008.

[5] E. H. Adelson and J. Y. Wang, "Single lens stereo with a plenoptic camera," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, no. 2, pp. 99–106, 1992.

[6] A. Levin, R. Fergus, F. Durand, and W. T. Freeman, "Image and depth from a conventional camera with a coded aperture," *ACM Transactions on Graphics*, vol. 26, no. 3, pp. 70:1–70:9, 2007.

[7] D. L. Marks, D. S. Kittle, H. S. Son, S. H. Youn, S. D. Feller, J. Kim, D. J. Brady, D. R. Golish, E. M. Vera, M. E. Gehm, R. A. Stack, E. J. Tremblay, and J. E. Ford, "Gigapixel imaging with the AWARE multiscale camera," *Optics and Photonics News*, vol. 23, no. 12, p. 31, 2012.

[8] D. L. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

 ISBN: 978-1-61208-296-7

[9] M. F. Duarte, M. A. Davenport, D. Takhar, J. N. Laska, T. Sun, K. F. Kelly, and R. G. Baraniuk, "Single-pixel imaging via compressive sampling," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 83–91, 2008.

[10] J. M. Kahn, R. H. Katz, and K. S. J. Pister, "Next century challenges: Mobile networking for 'Smart Dust'," in *Proceedings of the 5th annual ACM/IEEE International Conference on Mobile Computing and Networking (MobiComm 99)*, 1999, pp. 271–278.

[11] P. R. Gill, C. Lee, D.-G. Lee, A. Wang, and A. Molnar, "A microscale camera using direct Fourier-domain scene capture," *Optics Letters*, vol. 36, no. 15, pp. 2949–2951, 2011.

[12] P. R. Gill, C. Lee, S. Sivaramakrishnan, and A. Molnar, "Robustness of planar Fourier capture arrays to colour changes and lost pixels," *Journal of Instrumentation*, vol. 7, pp. C01–61, 2012.

[13] A. Wang and A. Molnar, "A light-field image sensor in 180 nm CMOS," *IEEE Journal of Solid-State Circuits*, vol. 47, no. 1, pp. 257–271, 2012.

[14] A. Wang, P. R. Gill, and A. Molnar, "Light field image sensors based on the Talbot effect," *Applied Optics*, vol. 48, no. 31, pp. 5897–5905, 2009.

[15] S. Sivaramakrishnan, A. Wang, P. R. Gill, and A. Molnar, "Enhanced angle sensitive pixels for light field imaging," in *IEEE International Electron Devices Meeting (IEDM)*, 2011, pp. 8.6.1–8.6.4.

[16] C. Peroz, S. Dhuey, A. Goltsov, M. Volger, B. Harteneck, I. Ivonin, A. Bugrov, S. Cabrini, S. Babin, and V. Yankov, "Digital spectrometer-on-chip fabricated by step and repeat nanoimprint lithography on pre-spin coated films," *Microelectronic Engineering*, vol. 88, no. 8, pp. 2092–2095, 2011.

[17] P. R. Gill and D. G. Stork, "Lensless ultra-miniature imagers using odd-symmetry spiral phase gratings," 2013, optical Society of America Sensors Congress.

[18] R. L. Morrison, "Symmetries that simplify the design of spot array phase gratings," *Journal of the Optical Society of America A*, vol. 9, no. 3, pp. 464–471, 1992.

[19] P. R. Gill, "Odd-symmetry phase gratings produce optical nulls uniquely insensitive to wavelength and depth."

[20] ——, "Odd-symmetry phase gratings produce optical nulls uniquely insensitive to wavelength and depth," 2013, submitted for publication.

[21] D. G. Manolakis, V. K. Ingle, and S. M. Kogon, *Statistical and adaptive signal processing: Spectral estimation, signal modeling, adaptive filtering and array processing*. Norwood, MA: Artech, 2005.

[22] D. A. Fish, A. M. Brinicombe, E. R. Pike, and J. G. Walker, "Blind deconvolution by means of the Richardson-Lucy algorithm," *Journal of the Optical Society of America A*, vol. 12, no. 1, pp. 58–65, 1995.

[23] H. Chaouchi, Ed., *The Internet of Things: Connecting objects*. New York, NY: Wiley, 2010.

[24] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, 2nd ed. New York, NY: Wiley, 2001.

 ISBN: 978-1-61208-296-7

Atmospheric Icing Sensors – An Insight

Umair Najeeb Mughal and Muhammad Shakeel Virk

Department Of Technology, Narvik University College,
Narvik, Norway
Email : unm@hin.no

Abstract **– To design a new atmospheric icing sensor, it is important to understand the existing atmospheric ice detection/measurement techniques. It is found that atmospheric icing sensors, which are capable of delivering maximum information are based upon direct measurement of the physical properties (electrical or mechanical) of atmospheric ice. Most of the ice detection methods are not able to deliver combined information about atmospheric ice type, its thickness and its rate simultaneously (very important factors for ice mitigation and design purposes). Based upon this study, it is found that sensors based upon the application of complex relative permittivity variations for different types of atmospheric ice at different ambient conditions offer good potential to detect atmospheric ice type, thickness and rate with minimum loading errors (effects of freezing ambient environment) due to their capacitive nature. This indicates that sensors based on capacitive and impedance measurement techniques are mostly suitable for the development of robust ice measurement sensors.**

Keywords **- Atmospheric Ice; Capacitive; Dielectric; Direct Measurement; Microwave; Resonance; Ultrasonic.**

I. INTRODUCTION

Due to increasing trend of activities in polar regions, human and inventory hazards are also increasing. Hence, development of robust techniques to detect and measure atmospheric ice accretion on structures is a very critical and challenging task. There is no ice sensor commercially available that can detect and measure all important icing parameters such as: icing rate, load and type under any icing conditions simultaneously [1]. Currently, most ice detectors available are capable of measuring either one or two phenomena such as detecting ice and indicating the rate of icing only. Ice sensors can be integrated with anti/deicing systems; therefore it is important for these to deliver sufficient information to be able to operate ice mitigation devices effectively. Power requirements for the removal of snow and ice is different hence to distinguish between snow and ice can be considered to be a limiting factor for de-icing system, because most devices for the removal of snow are normally ineffective for the efficient removal of ice or hard packed snow [2]. Also, measuring an icing event bounds a set of requirements, which include the ability of a sensor to detect icing with high sensitivity without influencing the measured quantity due to the ambient conditions, this is defined as the loading error. These atmospheric ic measurement techniques are categorized into indirect and direct methods [3]:

Indirect methods of ice detection involve measuring weather conditions such as relative humidity, wind velocity, temperature etc. that lead to the prediction of an upcoming icing event [4]. Empirical or deterministic models are used in indirect methods to determine when icing will occur.

The direct methods of ice detection are based on the principle of detecting property changes caused by accretion such as mass, dielectric constants, conductivities, or inductance. This paper briefly describes the working principle of different available direct measurement techniques (categorized below) of atmospheric ice on structures.

i. Capacitive based techniques,
ii. Microwave based techniques,
iii. Impedance based techniques,
iv. Ultrasonic based techniques,
v. Load cell based techniques,
vi. Infrared based techniques,
vii. Resonance based techniques.

II. ATMOSPHERIC ICE PARAMETERS MEASUREMENTS USING DIRECT MEASUREMENT TECHNIQUES

A. Capacitive based techniques

Capacitive ice sensors generate an electric field to detect the presence of dielectric materials. An electric field radiates outward around the probe and a dielectric material in close proximity of the field affects the measured capacitance, Mughal et. al. [5]. This attribute enables non-invasive measurements. In Sihvola et. al. [6], the results indicate that the complex dielectric constant is practically independent of the structure of snow. It is also mentioned that for dry snow, the dielectric constant is determined by the density and for wet snow, the imaginary part and the increase of the real part due to liquid water have the same volumetric wetness dependence. The static dielectric constants ε_{rs} of both polycrystalline and single crystals of ice have been carefully determined by Auty and Cole [7]. Also, application electrical properties to the measurement of ice thickness, temperature, crystal orientations are presented in Evanes [8]. Weinstein [9] and Jarvinen [10] proposed two different capacitive based ice detection methods, which are discussed in the following sections. Sensing methods proposed by Gerardi [11, 12] also use the same principle as proposed by Weinstein [9].

 ISBN: 978-1-61208-296-7

Capacitive ice sensor by Weinstein

The ice sensor proposed by Weinstein [9] as given in Fig. 1, can be used for the determination of the thickness of ice (22) on the outer surface (12) of an object independent of temperature and the composition of the ice (22). First capacitive gauge (16), second capacitive gauge (18), and the temperature gauge (20) are embedded in embedding material (14) located within a hollow portion of outer surface (12). First capacitive gauge (16), second capacitive gauge (18), and temperature gauge (20) are respectively connected to first capacitance measurement circuit (24), second capacitance measurement circuit (26), and temperature measuring circuit (28). The geometry of first and second capacitive gauge (16) and (18) is such that the ratio of voltage outputs of first and second capacitive gauge (24) and (26) is proportional to the thickness of ice (22), regardless of ice temperature or composition. This ratio is determined by offset and dividing circuit (29).

Both first (16) and second capacitive gauge (18) are made from thin conductors with a thickness of approximately 0.0254mm. A dual timer LM556 (42), used in an astable mode to generate 7μs pulses at any frequencies e.g 1.5kHz, to trigger a monostable timer (44). The timing capacitor of monostable timer (44) is (16). The output from the monostable timer is converted by the low pass filter (46) to produce an output DC signal, which is directly proportional to the capacitance of (16). Capacitance measuring circuits (24) and (26) are connected to offset the dividing circuit (29). The output voltage V_{out} of this offset and dividing circuit (29) for ice conditions is determined by the relation,

$$V_{out} = \frac{(V - V_o)_2}{(V - V_o)_1} \quad (1)$$

where V is the voltage output for the ice conditions and V_o is the initial voltage for no ice conditions. Subscripts (1) and (2) refer respectively to capactive measurements from first capacitance measuring circuit (24) and second capacitance measurement circuit (26). V_{out} is independent of both temperature and ice decomposition since, both effects results in identical scaling factors [13] resulting no changes in equation 14.

The variation of capacitance as a function of thickness is shown in Fig. 3. The output voltages from first and second capacitive measuring circuits (24) and (26) for various thickness of ice (22) formed on the outer surface (12) is shown. Curve (48) represents the output voltage V_1 from the first capacitance measuring circuit (24) at a fixed temperature and ice impurity level. Also curve (50) represents the voltage output V_2 from the second capacitance measuring circuit (26). At a fixed configuration, both curves will vary with temperature change or ice impurity change. Also it can be seen that both curves do not go to zero output voltage even when the ice thickness is zero.

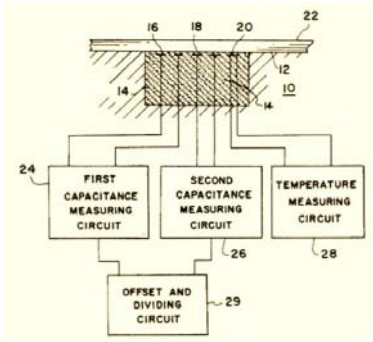


Figure 1. Construction of Weinstein Ice Sensor, Source[9]

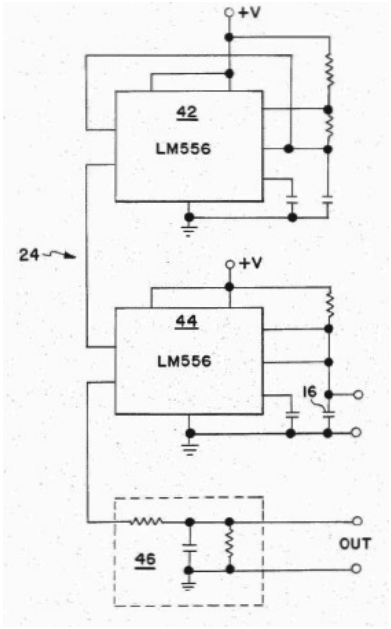


Figure 2. Electrical schematic of capacitance measuring circuit, Source[9]

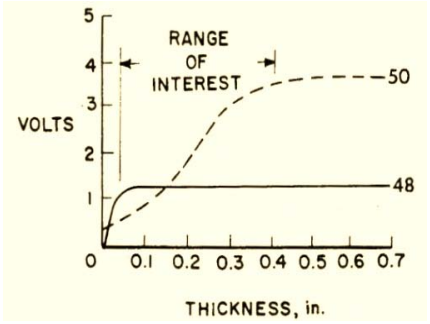


Figure 3. Ratio of capacitance gauge as a function of thickness, Source[9]

Capacitive ice sensor by Jarvinen

The ice sensor designed by Jarvinen [10] is capable of detecting the ice, its thickness and type. The dielectric properties of ice and snow show variation when an alternating current is supplied through it as mentioned by a review of Evans [8] and Stiles [14]. A lot of the dielectric measurements based upon Cole – Cole plot [15, 16] on wet snow, pure snow, granular snow and compact snow were conducted by Kuroiwa [17] and many others. In this sensor, Jarvinen [10] used the

 ISBN: 978-1-61208-296-7

method for detecting the presence and the accretion of ice by first measuring the properties of the contaminant layer overlying the ice sensor. The contaminant layer's temperature, thermal conductivity and variation of total impedance versus ice sensor electrical excitation frequency are measured. The complex dielectric property subsystem monitors the dielectric property locus in dielectric space as the excitation frequency is varied from near dc to higher frequencies and compares the measured results for magnitude and shape with laboratory property data taken at the same temperature and stored in the processor. It double check using external ice (based upon the complex dielectric measurements) sensor whether it is ice or rain water or deicing fluid or snow. If the measured results form a semicircular shaped locus of dielectric properties in complex dielectric space during the frequency scan and those measurements are also determined to be in agreement with on board stored laboratory ice data, ice is confirmed to be present. The presence of ice is also confirmed if a particular vector can be constructed from the measured data taken at a single preselected excitation frequency and found to have a vector angle in agreement with the vector angle from stored laboratory results taken at the same measurement conditions. In addition, complex dielectric measurement algorithms identify whether cracks, flaws or voids or increased electrical conductivity exist in the ice covering and sensor from their effects on the shape and size of the measured complex dielectric locus or from the length of the vector at the pre selected frequency. The presence of flaws, cracks or voids or enhanced electrical conductivity are determined from the values for the low frequency and high frequency intercepts and the value for diameter of the complex dielectric locus if these values are found to differ from those calculated for ice based on stored ice data. These differences, if found to exist, are used to correct the initially chosen ice thickness value based on the assumption of normal ice: ice with no flaws, cracks or voids or higher electrical conductivity, see Fig. 5. For more details on the mathematical principle of this type of sensing technique see Mugal et. al. [16].

Pure glaze ice in the temperature range from *0* to *-40^{o}C* has a thermal conductivity value in the range *2.4* to *2.6 W/mK*, rain water slightly above *0^{o}C* has a value of *0.6 W/mK* in the same units, air *0.023 W/mK* and a *50/50* mixture of deicing fluid is *0.41 W/mK*. The thermal conductivity of rime ice (density *0.38 gm/cm^3*) have a thermal conductivity of *0.4 W/mK*. Thus the presence of glaze ice is easily determined by the substantial difference in thermal conductivity between it and all other possible contaminants. For more details follow [13]. Dielectric values are used in addition to discriminate between rain water, deicing fluid and low density rime ice.

Also, this sensor uses *AD5933* high precision impedance converter system that combines an on board frequency generator with a *12 bit, 1 MSPS, ADC*. The frequency generator allows an external complex impedance to be excited with a known frequency. The response signal from the impedance is sampled by the on board *ADC* and a DFT is processed by an on board DSP engine. The DFT algorithm returns a real and imaginary data word at each output frequency. *AD5933* chip measures total impedance magnitudes over the range from *100Ω* to *10MΩ* with a total system accuracy of *0.5%*. The sweep frequency range of the chip normally covers the range from *1kHz* to *100kHz*. By adding a clock dividing circuit we can lower the range to *10Hz* to *20KHz*. For this sensor this range is *40Hz* to *40KHz*.

From Impedance to Dielectric : It is explained in [18] that ice can may be replaced by an equivalent circuit as shown in Fig. 4. The total complex impedance Z of the circuit is given by,

$$\frac{1}{Z} = \frac{1}{R_0} + \frac{1}{R_1 + \frac{1}{j\omega C_1}} + j\omega C_\infty \qquad (2).$$

Now if the ice is characterized by a capacitance *C*, then

$$\frac{1}{Z_c} = j\omega C = \frac{j\omega A \varepsilon_o \varepsilon_r}{L} \qquad (3)$$

where *L* is thickness of ice, *A* is the surface area of block of ice and ε_o is the permittivity of free space. Also, we have

$$\varepsilon_{rs} = \frac{L}{\varepsilon_o A}\left(C_1 + C_\infty\right) \quad \& \quad \varepsilon_{r\infty} = \frac{LC_\infty}{\varepsilon_o A} \qquad (4)$$

Eq. (4) when substituted in Cole-Cole equation we get,

$$\varepsilon_r = \frac{\varepsilon_{rs} - \varepsilon_{r\infty}}{1 + \left(j\omega\tau_0\right)^{1-\alpha}} + \varepsilon_{r\infty} = \frac{LC_1}{\varepsilon_o A\left(1 + j\omega\tau\right)} + \frac{LC_\infty}{\varepsilon_o A} \qquad (5)$$

where $\tau = R_1C_1$ is the dielectric relaxation time of the circuit that is time taken for the voltage across the ice block to reach $1\text{-}e^{-1} = 0.63$ of its final value when a step voltage is applied.

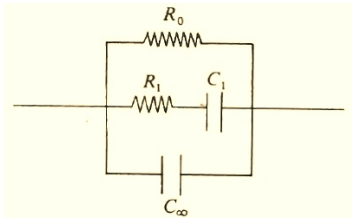


Figure 4. Equivalent circuit for ice [18]

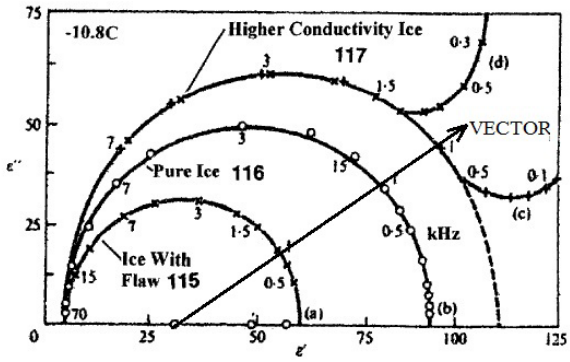


Figure 5. Cole Cole plot for pure ice, ice with flaw & ice with higher conductivity value, Source [10]

B. Ultrasonic based techniques

Ultrasonic ice sensors consist of two transducer elements, where one element generates ultrasonic vectors, which are detected by the second element. By measuring the attenuation levels, icing between the two elements can be detected. Like capacitive sensors, ultrasonic sensors need low power, low cost and are directionally sensitive. Luukkala [19] and Watkins [20] proposed two different ultrasonic based ice sensing techniques, which are discussed in the following sections.

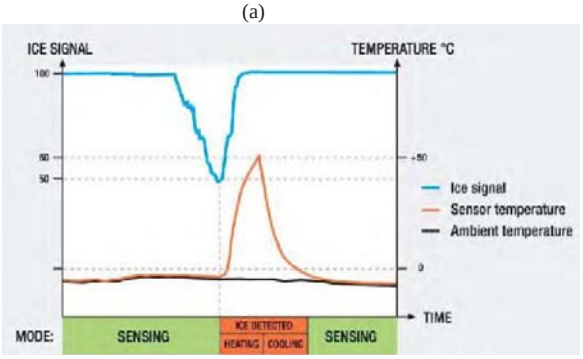
Ultrasonic ice sensor by Luukkala

In the ice sensor designed by Luukkala [19] , a mechanical ultrasonic signal is transmitted along a thin thread or strip at one end and the intensity of the ultrasonic signal having passed through the thread is measured at the other end. If the thread is covered with a water layer, the ultrasound will not be attenuated, however if the water freezes, the ultrasound cannot propagate in the thread, but will be abruptly attenuated. If the thread is covered with a sludge, the ultrasound will be somewhat attenuated to a kind of intermediate level at which detection of sludge is also possible. A viscosity difference exists between ice and water, and thus the intensity of the ultrasound having passed through the thread will also be different.

These ultrasonic sensors are comprised of a measuring transducer, a thread-like or tape-like acoustic waveguide having an ultrasonic transmitter at one end, an ultrasonic receiver at other end, and a device which comprises electronic components for measuring the intensity and the attenuation of an ultrasonic pulse having passed through the transducer thread in the case of ice formation. Electric resistance of the thread is measured simultaneously with the measurement of the attenuation of the ultrasound. With the help of measured resistance, the amount of required heating for the thread can be provided optionally to melt the surrounding ice, whereby the ultrasound intensity resumes its initial level.

This technique has been practically utilized in LID-3300IP [21] and are primarily aimed to detect ice on wind turbines, see Fig. 6a. During icing conditions the ultrasonic signal amplitude will start to decrease and the ice alarm will make the turbine stop to start the blade heating. Also, just at the right time when the ice is detected the sensor start heating itself to melt the detected ice. After a set delay, the alarm will go off and turbine will start functioning again, see Fig. 6b[22]. This sensor have been approved by the Technical Research Centre of Finland (VTT), and more than 2,000 units have already been supplied to the largest wind turbine manufacturers LID ice detectors are in use at airports and weather stations also. This ice detector is based on longitudinal wire waves [4]. It is reasonably east to adjust the parameters of the device to correspond with different icing climates. However, Labko's different versions have suffered from snow induced icing conditions and inability to melt all the ice.

(a)



(b)

Figure 6. Labcotec Ice Detector, Source [22] (a) Labcotec Ice Detector installed on the nacelle of wind turbine, (b). Working cycle of detector

Ultrasonic ice sensor by Watkins

The ultrasonic sensor designed by Watkins [20] consists of two ultrasonic transducers. The first transducer is energized to cause propagation of ultrasonic waves through a portion of a solid metal sheet along its predominant component parallel to the surface of the sheet and detecting the waves by means of the first ultrasonic transducer, while measuring the amplitude of the waves received by the second transducer, on the surface. If a layer of ice forms on the surface, the amplitude and intensity of the waves detected by the second transducer will decrease [23]. The waves may be horizontally polarized guided shear waves, or the waves may be a mode of lamb wave [24] whose predominant component is horizontal. The transducers may be attached to the opposite surface of the portion of the sheet to which an ice layer may develop, and may be generated and received by piezoelectric or electromagnetic means.

In Fig. 7a, the design of Watkins's sensor to detect a layer of ice on a thin solid sheet is shown. Two ultrasonic transducers are attached to the sheet; the first transducer is adapted to cause propagation of ultrasonic waves through the thin solid sheet having their predominant component parallel to the surface of the sheet. Such waves will dissipate energy into an icing layer adhering to the surface, but not into air or liquid water, in the case where liquid water is present. The second ultrasonic transducer is adapted to detect the waves propagating in the sheet and to give a signal representative of their amplitude. A discriminator responsive to the signal

 ISBN: 978-1-61208-296-7

detects the presence of an ice layer on the surface. In Fig. 7b it is shown that the transducer comprises six transducer strips parallel to each other extending perpendicular to the plane of the figure and spaced apart at a distance equal to the wavelength of ultrasonic shear waves in the sheet, held in a solid matrix. Each strip is of length 30 mm and is bonded to the surface by an electrically conducting layer and is made of piezo electric material and has contacts on opposing top and bottom surfaces by means of which it can be excited into vibrations parallel to its length. These vibrations are in phase with the other strips, and cause horizontally polarized shear waves to propagate in the sheet.

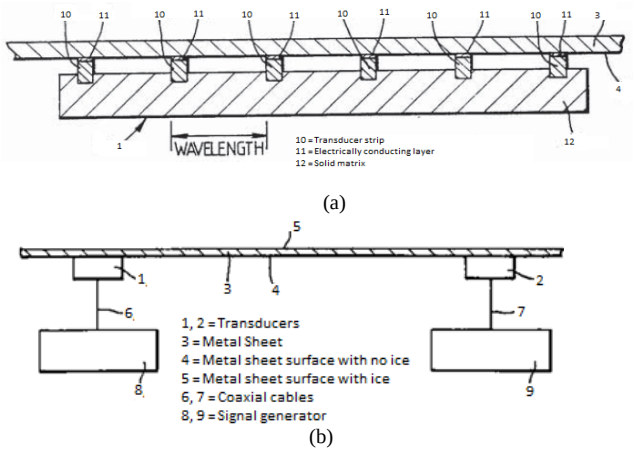


Figure 7. Working Principle of Watkins's Icing Sensor ; Source [20]

B. *Resonance based technique*

In this ice sensing technique an element is allowed to vibrate at its resonance frequency. Due to the deposition of ice or any other mass on the vibrating element, its frequency spectrum varies. The variation in primarily fundamental/resonance frequency will then deliver information about the icing event. This type of ice sensor is a single point ice detection device, which delivers information about the presence of ice and cannot distinguish between the types of icing. Cronin [25], Kossmann [26] proposed two different resonance based ice sensor designs, which are discussed in the following sections.

Resonance ice sensor by Cronin

This ice sensor design [25] uses the principle of magnetostriction, which is defined as the ability of ferromagnetic materials to change dimensions under the influence of a fluctuating magnetic field (see Fig. 8). These magnetostrictive sensors vibrate ultrasonically at a set resonant frequencies. As ice accretes on the probe, the vibrational frequency decreases. At a specified frequency shift, usually a decreasing frequency, which is related to the ice mass on the probe, an output signal is generated and probe heaters are energized to remove the ice. The heater is then de-energized and when the probe cools down, begins to accrete ice again if the icing event was still persistent. This sensing method has been practically utilized in [27, 28]

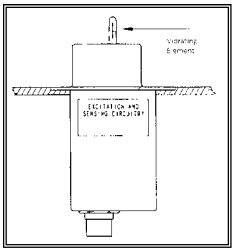


Figure 8. Cronin resonance based ice sensor; Source [25]

Resonance ice sensor by Koosmann

The ice sensing technique proposed by Koosmann [26] was not novel but presented a further improvement on the method proposed by Werner et. al. [29]. This technique is different from the sensing technique proposed by May [30] and Roth [31] who used a diaphragm as a vibrating element. In the technique proposed by Koosmann [26] a vibrating element in a tube vibrates along the longitudinal axis of the tube. This tube is driven by an excitation coil at its natural frequency and is sealed by a diaphragm, which has a surface exposed to an air stream in which icing is to be sensed. The exposed diaphragm surface is deflectable during vibration of the tube at a flexible support portion of the diaphragm. As ice accumulates on the exposed surface the natural frequency of the cylindrical section changes and is sensed to determine that ice is accumulating. The diaphragm is of low mass and small size so that the stiffness of small amounts of ice can significantly change the spring constant of the flexible support. The diaphragm is also shaped to conform to adjacent aerodynamic surfaces.

C. *Microwave based technique*

In this technique microwave energy is allowed to pass through a substance using a waveguide. The amount of reflected energy can then be used to deliver information about the icing event and the amount of ice. Overall [32] and Magenheim [33] proposed two different microwave based ice sensors, which are discussed in the following sections.

Microwave ice sensor by Overall

The icing sensor designed by Overall [32] can detect the presence of ice or water on a road or other surfaces. In this detector, microwaves are directed via a waveguide to the underside of a window substantially transparent to the microwaves and installed substantially flush with the surface to be monitored. The presence and thickness, within reasonable limits, of any coating of ice or water present on the surface of the window is determined by measuring the amount of microwave energy reflected by the window. The reflected energy can be used to trigger a signal device such as a warning sign or to activate ice melting means.

 ISBN: 978-1-61208-296-7

In Fig. 9 it is shown the block diagram of the basic elements of this sensor design, which serves to define the principle of operation. The microwave generator is coupled with the ferrite isolator, which functions to protect the generator from the reflected waves and to keep the power level and frequency of the generated signal constant in spite of variations in the reflected signal and drastic load changes. The waveguide couples the ferrite isolator to the window, which can be constructed of any dielectric element that is substantially transparent to microwave energy, as for example polymers of tetrafloroethylene, copolymers of hexafloropropylene and tetrafloroethylene, polyethylene. The microwave energy reflected at the window travels back along the hollow waveguide and meets a directional coupler, which bleeds from the main guide a small proportion of the reflected energy. A crystal rectifier positioned in the directional coupler is actuated by the reflected energy and the output from the crystal is amplified and displayed on an amplifier to quantitatively show the level of the reflected energy.

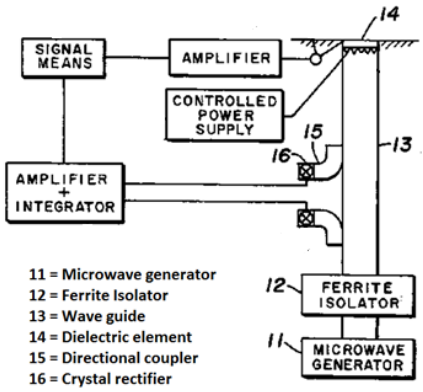


Figure 9. Microwave ice detector by Overal; Source [32]

Microwave ice sensor by Magenheim

In this design [33], microwave electromagnetic energy (typical frequency falls 2k – 20k MHz) is transmitted into a surface of ice, and the reflection or impedance characteristic of the ice layer is monitored to determine the presence and amount of ice. This ice sensor includes a device to generate microwave electromagnetic energy, transmitting the microwave energy in to a surface layer of ice. The microwave energy reflected from the ice layer is used to determine the presence and relative amount of ice in the layer. It also includes a permanent surface waveguide of such a thickness as to allow the propagation of microwave energy even when the ice layer is extremely thin. The design of this ice detector includes a short circuit element at the end of the surface waveguide remote from the end into which the microwave energy is transmitted. The short circuit element results in a partial or total reflection of the energy back along the surface waveguide and reflected energy being detected by the reflecto-meter means. Since ice typically contains many impurities including a significant amount of unfrozen water, it will present a dielectric medium for the transmission of the microwave energy, and much less energy is reflected out of the waveguide when a layer of ice is present.

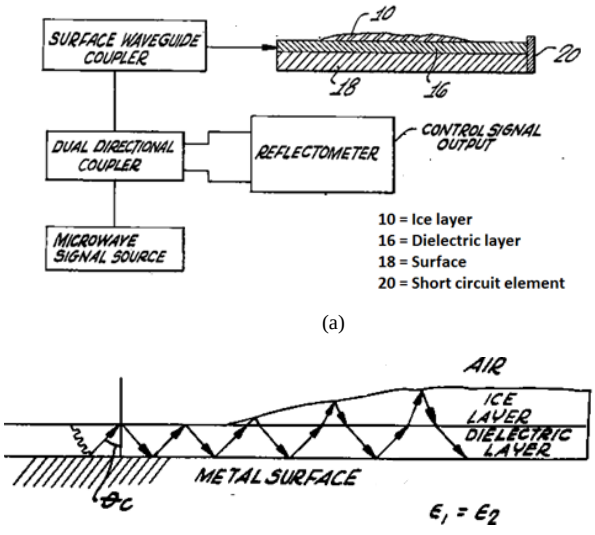


Figure 10. Microwave ice detector by Magenheim; Source [33], (a) Transmittance of microwave energy in iced surface, (b) Working principle

D. Impedance based technique

The impedance based ice sensing technique is very similar to the capacitive measurement technique. The only difference between the two is the source of information delivered, which is current in the former and voltage in the later. The importance of the impedance measurement system is that we can avoid saturation of the electrodes in this technique and we can increase our data points.

Impedance ice sensor by Seegmiller

The design of Seegmiller [34] is capable of delivering parameters such as the thickness of ice and the ice load for a relatively wide surface. This ice sensor is comprised of inductive ice sensing electrodes, temperatures sensors, a frequency generator, a voltage detector, resistance bridge and a processing unit. The inductive ice sensing electrodes are flush mounted on the surface of interest and comprise a transmitting electrode (at least one receiving electrode). These electrodes are insulated to avoid a spurious reading caused by conductive substance or electrolytes. The coefficient of coupling between the transmitting and receiving electrodes is determined by at least two factors, the first being the predetermined geometry and spacing between the electrodes and the second being the inductive coupling susceptibility of the ice in the general region of the spaced apart electrodes. This susceptibility is indicative of the presence and thickness of ice. The device for measuring temperatures is flush mounted onto the surface of interest. The frequency generator supplies an excitation signal and has means for being connected to the transmitting electrode of the inductive ice sensing array. The voltage detector has means for being connected to the receiving electrode of the inductive ice sensing array and detects a proportion of the supplied excitation signal to the transmitting electrode as a function of the coefficient of coupling between the transmitting and the

 ISBN: 978-1-61208-296-7

receiving electrode. The voltage detector device generates an output signal representative of the thickness of adhering ice, frost or water substance in the general region between the spaced apart electrodes.

E. Infrared based techniques

This technique is based on the absorption and reflection of active infrared light by the ice. By measuring the amount of reflected light, one can determine if ice is present at the surface.

HoloOptic ice sensors

HoloOptics sensor has the potential to detect the presence of any type of ice or snow. The working element comprises an infrared emitter together with a single head, two head or four head receivers, a photo detector and a probe. The operating range of this sensor lies in NIR range (0.88 μm to 0.92 μm), [35]. An icing event is recorded if more than 95% of the probe is covered with a 5nm thick layer of glaze ice or a 9nm thick layer of other types of ice. Once icing is detected, the probe internal heating system is activated to melt the accreted ice. The time it takes to deice is dependent solely on the icing rate if sufficient amount of heating power is provided. The time lapse between icing events is used to determine the icing rate [36].

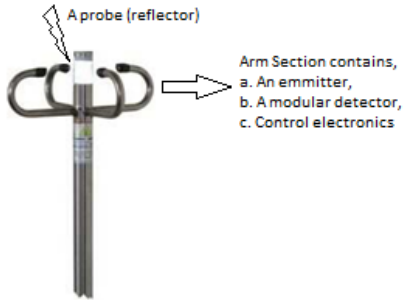


Figure 11. T44 Holoptic ice sensor; source [36]

F. Load cell based techniques

In the international standard ISO 12494 [37], a standard way of measuring ice accretion is suggested. In this method it is required to measure the load of atmospheric ice on a steel rod that is 0.5 meter high (1 meter if heavy icing is expected) and has the diameter of 30mm. The rod must be freely rotating or forced to rotate at a constant speed. When ice accretes on the steel rod, aerodynamic drag will cause it to rotate in the case of free rotating icing load monitor, always facing the least amount of the iced part towards the wind. This doubling of length of the steel rod due to heavy icing is primarily aimed to uniform the drag distribution along its profile. By measuring the weight of the iced steel rod with the help of load cells, the amount of ice that has accreted can be determined. Two ice sensors (ice monitor & ice meter) have been developed on the basis of this technique. The following sections describes these two sensors:

The IceMonitor

The IceMonitor® measures the mass of accumulated ice gravimetrically. The working element is a freely rotating steel pipe resting on a rod placed on a load cell. As ice accretes on the freely rotating steel pipe, the ice load is weighed by the load cell. The Ice Monitor is manufactured by SAAB Technologies and was initially developed for power line surveillance systems. It can measure the rate because the readings from the load cell are recorded with time. The Ice-Monitor® is not able to detect ice over a wide area cannot distinguish between the two types of in-cloud icing.

Ice Meter®

The ice meter was developed by the Institute of Atmospheric Physics, Prague, Czech Republic. It measures the mass of icing accumulated on the surface of the collector. It has a horizontal rod, which is coupled with a cylindrical collector to the tensometric. The cylinder is orientated vertically in order to eliminate the detection of wet snow as much as possible but the orientation of this cylinder can be changed to horizontal, if required. In this sensor, the mass of accumulated ice is measured by means of a tensometric bridge (strain gage load sensor) the output of which is tied to the precise A/D converter. The digital signal is preprocessed by a micro-controller, which assigns the time and stores the data into the device memory. In order to prevent the freezing of the horizontal rod, which couples the cylindrical collector to the tensometric, which is located together with the electronics in the housing. The passage through the housing may be heated depending on the passage temperature. A test electromechanical impulse is applied each hour to verify the free force transition to the tensometric, and thus to check whether the acquired data are reliable or not [38].

III. PROS/CONS OF AVAILABLE SENSORS

A. Measurements in Finland

The icing climate of the site has been characterized as site class A: an elevated site inland in northern Europe with harsh and frequent icing climate. The Labko LID-3503 and the Rosemount 0872J were installed at the site in the winter 2001/02. Both ice detectors indicate the presence of icing conditions. According to the measurements performed with the LID-3503 ice detector it was noted that the sensor is inadequate for icing measurements in extreme conditions. Also, the conventional humidity instruments measure incorrectly in icing conditions [4].

 ISBN: 978-1-61208-296-7

B. Measurements in Czech Republic

The icemeter has been operated on the Milesovka peak from 2000. Although icemeters installed in the site measured correctly most of the time, but there were also some time periods when the instruments gave obviously wrong negative values. These times can usually be associated with the periods when the horizontal rod coupling the tensometric with the vertical collector became icebound to the instrument housing. Thus there was no free force transition, which can be identified by not observing a proper electromechanical pulse in the data. The following changes are expected to be made in icemeter for its better performance [4],

i. Possibility to build an instrument with a rotating collector.
ii. More focus on the sensors that measure accumulated icing.

C. Measurements in Finland

Rosemount 0872J (prototype) at the FMI's test est. The test station is located in northern Finland on the top of Luosto fell (500m asl, N 67 08', E 26 54'). The Luosto test sites represent an elevated site inland with harsh and frequent icing climate. According to the performance of the instrument used fo rice detections were not entirely reliable and difference were found between the performances of ice detectors. The ice detectors are to some extent insensitive to icing under heavy icing conditions. It was possible to record more or less accurately the start and ending of icing periods but not the accretions rate or type of the icing. The Rosemount sensor has yielded fairly good measurements at Luosto and detected the presence of icing conditions. In soft icing conditions ice accretion may exist on the sensor probe during short periods of time especially in the beginning of the icing event but the sensor does not detect ice. This sensor is fairly adequate for icing measurements and it operated better than the other used instruments. Nevertheless, it cannot guarantee accurate measurements in all icing conditions especially in soft icing conditions [4].

IV. CONCLUSION AND SUGGESTIONS FOR FUTURE WORK

A. Conclusions/Findings

Based upon this review study of design and working principle of various ice measurement techniques, it is found/observed that sensors based upon capacitive and impedance techniques offer reasonable potential to detect important parameters pertaining to ice accretion such as ice type, ice thickness and icing rate. There is no extensive experimental and laboratory based study on the performance evaluation of all these sensing techniques because most of the literature found is based on theoretical performance evaluation hence it is a design conjecture. It is possible to combine two or three different techniques together particularly capacitance, impedance and resonance in order to measure these parameters and empirically develop an analytical relationship for the measurement of the icing parameters including ice load, it is anticipated to produce a more robust icing measurement device. In [3], Homola et. al. have described the unsuitable methods and suitable methods for detecting ice on wind turbines. He has also created a room for further investigations on the capacitive, inductance and impedance type of measurement techniques, which according to him appear the most promising sensing technique/method, which is also based upon conjecture without any detailed design study. During this study it is also observed that the sensor utilizing the natural electrical properties of atmospheric ice [14, 17] working on an efficient algorithm based upon the electromagnetism theory will be expectedly prove an efficient sensor and the practical example of this sensor is Jarvenin's total impedance and complex dielectric property icing detection sensor. If we adopt the physics of this sensor along with a rotating part then we may get a robust sensor. As the capacitive sensing technique, provide less loading errors than other techniques therefore, it may essentially be a part of new robust atmospheric icing sensor.

B. Future Work

Detailed performance evaluation based upon the laboratory experiments and on situ experiments of the different sensing techniques/methods is being developed. A prototype sensor based upon the dielectric potential of atmospheric ice is designed and being developed in the cold room chamber of NUC, which can be then optimized according to the area of application. Although capacitive based sensing method seems reasonable for the measurement of atmospheric ice type, rate and thickness yet it has some coupling issues due to the fringing effects (electric field lines try to go through the air and into the sensor head on the sides), which may lead to inefficient calculations. Hence, to avoid this, the same sensor can be then further improved to develop a robust atmospheric icing detection and analyzing scheme by utilizing other sensing techniques such as resonance, impedance, inductance etc., which are ought to be reasonable. Also a big challenge to build a robust sensor which can work in freezing domains, material selection alongwith humidity protection is very critical.

REFERENCES

[1] U.N. Mughal, M.S. Virk, and M.Y. Mustafa, "Electromagnetism based atmospheric ice sensing technique - A conceptual review", International Journal of Multiphysics, 2012. 6(4): pp. 341-353.
[2] C. D. Kuhmonen, "Ice removing machine", 1980.
[3] M. C. Homola, P.J. Nicklasson, and P.A. Sundsbo, "Ice sensors for wind turbines", Cold Regions Science and Technology, 2006. 46: pp. 125-131.
[4] S. Fikke, "Atmospheric icing on structures, measurement and data collection on icing, state of the art", 2006, MeteoSwiss. pp. 110.
[5] U. N. Mughal, M.S. Virk, and M.Y. Mustafa, "Atmospheric Icing Sensors - Capacitive Techniques", Sensors and Transducers, 2013. 18(Special Issue): pp. 24-32.

 ISBN: 978-1-61208-296-7

[6] A. Sihvola, E. Nyfors, and M. Tiuri, "Mixing formulae and experimental results for the dielectric constant of snow". Journal of Glaciology, 1985. 31(108).
[7] R. P. Auty and R.H. Cole, "Dielectric Properties of Ice and Solid D_2O", Journal of Chemical Physics, 1952. 20(8).
[8] S. Evans, "Dielectric properties of ice and snow - a review". Journal of Glaciology, 1965. 5: pp. 773-792.
[9] L. M. Weinstein, "Ice Sensor", 1988.
[10] P. O. Jarvinen, "Total impedience and complex dielectric property ice detection system", 2008.
[11] J. J. Gerardi, et al., "Apparatus for measuring ice distribution profiles", 1995.
[12] J. J. Gerardi, P.R. Dahl, and G.A. Hickman, "Piezoelectric Sensor", 1993.
[13] W. D. Coles, "Experimental determination of thermal conductivity of low density ice". Nace Technical Note 3143, 1954.
[14] W. H. Stiles, and F.T. Ulaby, "Dielectric properties of snow". Journal of Geophysical Research, 1981. 85(C2).
[15] R. H. Cole, and K.S. Cole, Journal of Chemical Physics, 1942(9).
[16] U. N. Mughal, M.S. Virk, and M.Y. Mustafa. "Dielectric Based Sensing of Atmospheric Ice", 38th International Conference on Application of Mathematics in Engineering and Economics. 2012. Sozopol, Bulgaria: In press.
[17] D. Kuroiwa, "The dielectric property of snow", Union Geodesique et Geophysique International. Association Internationale of Hdrologie Scientifique. Assemblee generale de Rome, , 1956. 4: pp. 52-63.
[18] P. V. Hobbs, "Ice Physics", 2010, Oxford University Press.
[19] M. Luukkala, "Detector for indicating ice formation on the wing of an aircraft", 1995.
[20] R. D. Watkins, et al., "Ice detector", 1986.
[21] http://www.labkotec.fi/en/products/ice_detector_for_wind_turbines/.
[22] http://windren.se/WW2011/55_Labkotec_Hautamaki_Martola_IceDetector.pdf.
[23] J. Wu, "Sensitivity of Lamb wave sensors in liquid sensing", IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control, 1996. 43(1): pp. 71-72.
[24] http://en.wikipedia.org/wiki/Lamb_waves.
[25] D. J. Cronin, D.G. Jackson, and D.G. Owens, "Ice detector configeration for improved ice detection at near freezing conditions", 2001.
[26] M. R. Koosmann, "Membrance type non intrusive ice detector", 1986.
[27] Goodrich Freezing rain sensor. Brochure: 4085 LIT 03/02,, 2002a.
[28] Goodrich Sensor Systems : Primary and Advisory Ice Detection Systems. Brochure 4079 LIT 03/02, 2002b.
[29] F. D. Werner and E.A. Grindheim, "Ice detector", 1967.
[30] M. H. May, "Condition responsive device", 1947.
[31] W. Roth, "Ice detector", 1966.
[32] W. W. Overall and W. Woods, "Ice detection apparatus employing microwave reflectance", 1974.
[33] B. Magenheim, "Microwave ice detector", 1977.
[34] H. L. Seegmillar, "Ice detector and deicing fluid effectiveness monitoring system", 1996.
[35] http://holooptics.utrymmet.com/Dokument/651.Userguide%20T40.004.En.pdf.
[36] http://holooptics.utrymmet.com/Dokument/651.Userguide%20T40.004.En.pdf.
[37] M. H. Foder, "ISO 12494 - Atmospheric icing on structures and how to use it", Proceedings of the Eleventh (2001) International Offshore and Polar Engineering Conference. 2001. Norway.
[38] http://www.ufa.cas.cz/html/upperatm/chum/namraza/Icingmeasczech2.pdf.

ACKNOWLEDGEMENTS

The work reported in this paper was partially funded by the Research Council of Norway, project no. 195153/160 and partially by the consortium of the project ColdTech- Sustainable Cold Climate Technology.

 ISBN: 978-1-61208-296-7

A Domain-Specific Platform for Research in Environmental Wireless Sensor Networks

Sebastian Bader, Matthias Krämer, Bengt Oelmann
Department of Information Technology and Media
Mid Sweden University
Sundsvall, Sweden
sebastian.bader@miun.se

***Abstract*—Wireless Sensor Networks have the ability to improve a multitude of existing application domains. These networks are built up from a number of sensor nodes with sensing, communication and processing capabilities and the performance of the networked system is defined by the performance of the node platform it is based on. In this paper, we present SENTIO-em, a hardware platform for research in the environmental monitoring application domain. Based on the application domain requirements, the architecture and implementation of SENTIO-em is optimized for environmental monitoring constraints, while it is sufficiently flexible to be reused for different applications within the domain. The architecture of the platform is presented and evaluated under both laboratory and different environmental conditions. The obtained results are compared to a number of existing node platforms, demonstrating that SENTIO-em provides high energy efficiency with increased processing performance, short state transition times, and low quiescent currents.**

***Keywords*-Sensor Node Platform, Environmental Monitoring, Domain-specific Design**

I. Introduction

Sensor nodes are the basic building blocks of any wireless sensor network. It is therefore of great importance that the platform on which the sensor node is based, supports the application requirements for the usage of the the intended system. Furthermore, the required tasks for the sensor node should be carried out as efficiently as possible, in order to prolong the lifetime of the system, to constrain the development time, as well as the system costs.

In final product solutions this will typically result in an application-specific hardware design, supporting the task execution without unnecessary overhead. In research applications, however, hardware platforms have to be reusable and hence make application-specific solutions inefficient. To allow for reusability, the platform has to be sufficiently flexible for it to be adjusted for requirements of different applications. Nevertheless, the system should perform well in each of the different use cases and thus, should not be hindered by its own flexibility.

A common approach to achieving reusability of sensor node hardware is the design of general-purpose hardware platforms, such as those presented in [1]–[4]. These platforms typically have no specific application in mind, but are designed to be usable as a prototyping platform for a wider area. On the downside, general-purpose nodes require more development time once they are used in specific applications. Additionally, they may not allow the performance of certain application tasks or are inefficient in their execution.

In this paper, we are introducing SENTIO-em, a hardware platform that is specifically intended for wireless sensor network applications in the environmental monitoring domain. By designing the sensor node platform in an application-domain-specific manner, we are able to include typical domain-specific requirements, while maintaining a flexible platform for operations in different applications within the domain. In this way, the platform becomes reusable without any loss in application focus or increase in overhead.

II. Related Work

The application domain of environmental monitoring presents requirements that are well matched by wireless sensor network technology. This has led to the adoption of this technology in a number of field-trial evaluations, such as those presented in [5]–[7], where a network of sensor nodes has been utilized as a measurement instrument. Despite their differences in the application constraints, a similar fundamental platform architecture remains.

Since the beginning of wireless sensor network research, a large number of node platforms have been introduced. As these represent too many different systems, only a subset of solutions are mentioned. A more detailed presentation of existing hardware platforms can be found in [8].

The systems proposed in [1]–[4] represent general-purpose research platforms, which are mainly targeted for wireless sensor network prototyping. While the general architecture of these nodes is similar, different device choices have been made for system implementation. The platforms contain processing, storage, communication and power units, but require sensors and other extension to be attached via an expansion interface.

A different architectural choice has been made in [9]. While still a general-purpose platform, the whole system can be constructed of a variety of building blocks and therefore suits different scenarios. Although this building

 ISBN: 978-1-61208-296-7

block approach allows for very flexible use of the different modules, it is not suited for fast development times, because, in the majority of cases, an application-specific design will be required in order to assemble the building blocks into a complete platform.

In order to provide platform access to a wider community, several companies have made sensor nodes commercially available. Examples of these platforms include Libelium's Waspmote, Zolertia's Z1 and the Shimmer node. The majority of these platforms, however, have only limited application focus in order to be usable by a wider community.

With the SENTIO-em platform we propose an application-domain-specific design. Instead of basing the design choices of the platform on general-purpose constraints or on the requirements of a single application, the design of the platform has been based on typical requirements found in the application domain of environmental monitoring. Thus, the platform remains flexible to variations between different applications in environmental monitoring, while showing a higher domain specific performance as opposed to completely general-purpose platforms.

III. System Design

After an analysis of typical characteristics in environmental monitoring wireless sensor networks, we describe how these characteristics have been taken into account in the design of SENTIO-em, and present the resulting hardware and software architectures of the platform. A final overview on the system's implementation is given at the end of the section in Table III.

A. Environmental Monitoring Requirements

Environmental monitoring is typically performed by collecting sample data in the natural environment [10, chap. 2]. As wireless sensor networks enable for new measurement instruments with high spatial and temporal resolution without the requirement of a fixed infrastructure, they have been used in a plethora of different environmental applications, ranging from the monitoring of glaciers in Iceland [11] to the observation of rain forests in Australia [12].

Environmental monitoring wireless sensor networks typically operate in a time-driven manner, which means that they are collecting data at preordained intervals. The collected data, or information extracted from it, is in the majority of systems transfered to a central gathering point, which leads to a simple *sample-and-send* operation for individual sensor nodes. Due to the monitoring of slowly changing parameters, such as temperature, humidity or gas concentrations, the resulting sampling rate of sensor nodes in environmental monitoring is rather low as well. The number of nodes within the network, on the other hand, is commonly large in order to cover a given terrain with a desired spatial resolution.

This has two effects. On the one hand, the low sampling rate poses only limited processing power demands on the computational unit within the sensor node. On the other hand, the possibility of a large number of sensor nodes within the system requires the individual sensor nodes to be of low cost and the maintenance requirements to be kept at a minimum in order to make the system utilization economically feasible. In order to minimize maintenance, the lifetime of the sensor nodes must be long. As the sensor nodes, however, typically operate on capacity-limited energy storage devices, the lifetime is bound by the energy capacity available. An increase in capacity, however, also leads to an increase in cost and size of the sensor node. Thus, energy efficient operation is the principal requirement for the successful application of wireless sensor networks in the environmental monitoring domain.

Furthermore, the target setting for environmental measurement systems are typically outdoor environments, which means that the final system has to withstand harsh weather conditions and operate reliably under varying ambient conditions. Because these conditions are difficult to reconstruct in laboratory environments, the development of an environmental wireless sensor network requires outdoor test deployments. As it is often necessary during the development stage to make changes to already deployed sensor nodes, the underlying platform should support easy deployment, reconfiguration and recovery.

Finally, due to the long system deployment periods and the desire for low per-node cost, platform overhead is a parameter to be kept at a minimum. With platform overhead we describe components, modules and functionalities, which are included in a node platform although they are not required for the intended task to be accomplished. Typical examples of these are functionalities that are added in order to simplify system development.

B. SENTIO-em Hardware

The SENTIO-em hardware architecture has been designed with the requirements, as previously mentioned, in mind. A block diagram of the overall system architecture is presented in Figure 1. The detailed description of the design concepts and implementation is divided into sub-categories, which are organized according to computation, communication, sensing, power, form-factor and interfaces.

1) Computation: As is the case with the majority of existing sensor node platforms, the computational unit of SENTIO-em is built around a microcontroller. Due to the low processing capabilities, which are generally necessary and the requirement for low energy consumption, earlier systems are based on simple 8-bit or 16-bit processors, such as the Atmel ATMega family and the series of MSP430 controllers. However, as semiconductor technology has advanced since the introduction of these platforms, more recent microcontroller cores offer higher processing power with

 ISBN: 978-1-61208-296-7

Table I
COMPARISON OF TYPICAL MICROCONTROLLER CHOICES FOR THE IMPLEMENTATION OF HARDWARE PLATFORMS IN SENSOR NETWORKS

	ATMega1281	**MSP430F1611**	**AT32UC3L**	**EFM32G280**
Manufacturer	Atmel	Texas Instruments	Atmel	Energy Micro
Architecture	8-bit	16-bit	32-bit	32-bit
Clock Frequency [MHz]	0–16	0–8	0–50	0–32
Flash [kB]	64–256	48	16–64	32–128
SRAM [kB]	8	10	8/16	8/16
Operating Voltage [volt]	2,7–5,5 (<8 MHz) 4,5–5,5 (>8 MHz)	1,8–3,6 (<4 MHz) 2,7–3,6 (>4 MHz)	1,62–3,6	1,8–3,8
Current Draw				
Active [mA MHz^{-1}]	1	0,5	0,2	0,18
Sleep [µA]	5–100	1,1–75	5–45	0,6–0,9
Off [nA]	250	100	9	20

Table II
COMPARISON OF CONSIDERED RADIO TRANSCEIVER CHOICES FOR THE IMPLEMENTATION OF SENTIO-EM

	CC1101	**SX1233**	**CC2520**	**XBEE 802.15.4**
Manufacturer	Texas Instruments	Semtech	Texas Instruments	Digi International
Frequency [MHz]	315/433/868/915	433/868/915	2400	2400
Max. Data Rate [kbps]	500	600	250	250
Sensitivity [dBm]	-116	-120	-98	-100
Output Power [dBm]	-30 to +12	-18 to +17	-18 to +5	-10 to +18
Current Draw				
TX [mA]	12–34	16–95	25–37	45–250
RX [mA]	14–17	16–17	18–25	50–55
Sleep [µA]	0,2	0,1	<1	<10
Implemented Protocols	none	none	low-level	low / high-level

similar or even lower power consumption, thus increasing the energy efficiency dramatically. For this platform an Energy Micro EFM32 controller was chosen, which is based on an ARM Cortex-M3 processor core. Table I provides a comparison of previously used low-power microcontrollers and the EFM32 controller on SENTIO-em. The parameter comparison shows that the EFM32 microcontroller outperforms the other listed processor choices in almost all regards.

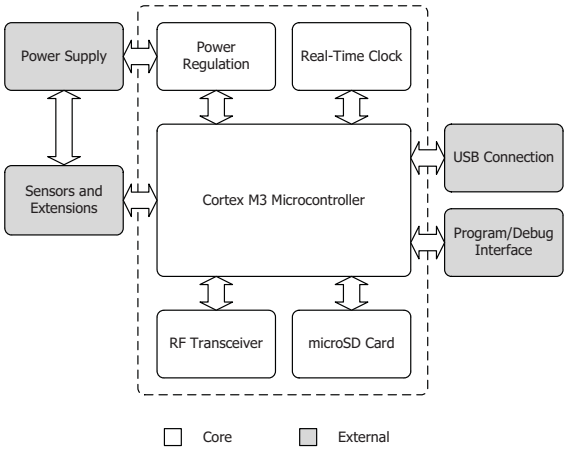


Figure 1. Block diagram of the SENTIO-em hardware architecture. The core node is shown in center, supported by application-specific extensions (left) and development functionalities (right).

The EFM32 controller is driven by a 32 MHz Q-MEMS crystal and has a connection to a Real-Time Clock (RTC) for accurate time-keeping. The RTC has an integrated crystal with temperature compensation, which improves the time-keeping accuracy in outdoor applications due to reduced environmental influences. In comparison to other controllers, the EFM32 stands out particularly because of its high energy efficiency.

2) Communication: The de-facto communication standard in research and commercial sensor networks are IEEE 802.15.4 based transceivers. These transceivers are available for both 868/915 MHz, as well as 2.4 GHz, both being part of the license-free ISM bands. Although this transceiver choice comes with some advantages, such as compliance with other systems, an existing community (i.e., available protocol solutions) and comfortable usage, its drawbacks are its limited flexibility and the considerable packet overhead. This plays a role, especially when payload sizes are low, as is typically the case in environmental monitoring applications. In environmental monitoring, moreover, lower frequency bands are preferable, as attenuation is reduced and hence longer communication ranges can be achieved with the same output power. On the other hand, low-frequency radio transceivers come without protocol implementations and thus require more development time.

Table II provides an overview of radio transceivers

 ISBN: 978-1-61208-296-7

(a) (b)

Figure 2. Pictures of the implemented SENTIO-em platform (a) mounted on the laboratory dock; (b) mounted in a protective enclosure. These two scenarios demonstrate the usage of the platform during development and deployment, respectively.

that have been considered for the implementation in the SENTIO-em platform. In addition to the standard transceiver chips, the XBee, a transceiver module with higher processing capability, has been listed. This module enables the implementation of high-level protocols (i.e., routing and networking) directly on the transceiver, but at the cost of higher overhead (e.g., energy consumption).

Due to the difficulty of predicting the exact communication requirements of individual applications, for the architecture of SENTIO-em a flexible solution has been chosen, which allows for the exchange of radio modules. The interface is based on the XBee transceiver module, which is available with different protocol implementations (e.g., 802.15.4, Zigbee or 802.11). Furthermore, low-level, low-frequency transceivers based on the Semtech SX1233 and TI CC1101 have been designed with the same interface in order to provide a flexible solution for custom protocols. By this means, it is possible to choose between general, easy to implement communication modules for rapid prototyping, or customized modules with optimized performance characteristics.

3) Sensing: Sensors are the modules which have the greatest variability between applications within the same application domain. While in environmental monitoring typical sensor types reoccur (e.g., temperature, humidity or barometric pressure), the amount, combination and requirements placed on the sensors can differ tremendously.

Due to this uncertainty in relation to sensor requirements and the targeted reusability within the application domain, SENTIO-em does not contain any sensors on its core module. Instead, sensors are connected to the platform via a 22-pin sensor interface, which provides a variety of microcontroller connection possibilities, including I2C, SPI, UART, as well as analog and digital I/O pins. In order to reduce overhead, the conversion to other interfaces, such as 4-20 mA or IEEE 1451, must take place on the sensor extension, if necessary.

4) Power Supply: In a similar manner to that for the sensor interface, a power interface has been implemented to allow for the usage of different power sources. While it is the case that during development platforms are usually powered from the grid, a laboratory power supply or an USB port, during deployments energy is typically taken from batteries or ambient energy sources.

The power interface provides a 16-pin connection to the designated power source module, which includes, in addition to the power path, communication and monitoring capabilities. The output from the power supply unit should lie between 3 V and 5.5 V, but can be unregulated as a power regulation unit is part of the core platform.

5) Form-Factor: Although the form-factor of a sensor node platform is typically not its first design criteria, it has a major influence on usability, size and cost of implemented systems. Initially, platforms were supposed to be tiny and inexpensive, but more recently usability has gained a greater focus. In the physical design of SENTIO-em we have attempted to incorporate the usability, in particular in relation to the environmental monitoring constraints as presented in Section III-A. While previous solutions have demonstrated form-factors that have enabled easy development by creating a USB-stick structure [2], or providing extreme reconfigurability from building blocks [9], in this platform both laboratory and deployment constraints were considered.

In the resulting physical implementation, modules that are only used in the laboratory development of the system are divided from the general platform core. As a result, a docking station-like board has been designed, whereon the SENTIO-em platform can be mounted during development. Because this laboratory docking solution is usually not deployed with the system, it is not restricted by size. Furthermore, as not every node platform requires its own docking station, price is also not a major concern. As a result, extra functionality can be provided for laboratory development, which, in a normal setting, would be omitted due to the hardware overhead. Moreover, the size and cost of the actual node platform can be reduced, as modules that are only required during development can be removed. Figure 2a depicts such a laboratory docking solution, which in this

 ISBN: 978-1-61208-296-7

Table III
OVERVIEW OF THE SENTIO-EM PLATFORM IMPLEMENTATION

Microcontroller	Energy Micro EFM32	
Clock Frequency	32 MHz	
RAM/Flash	16/128 KB	
Local Storage	SD card	
Time Keeping	Temp. compensated RTC	
Radio	CC1101/SX1233	XBee 802.15.4
Frequency	433 MHz	2.4 GHz
Data Rate	500/600 kbps	250 kbps
Development	USB, ARM 20	
Extension	Sensor, Power, Debug (72-pin)	
Size	75 mm × 49 mm	
Sleep-mode Consumption	<2 µA	

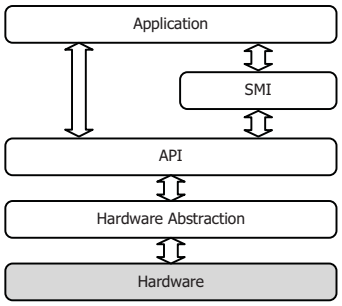


Figure 3. Layered software organization from hardware to application for SENTIO-em. The figure shows all available software layers, which do not have to be implemented in every application

case includes a USB interface, programming connection, a complete debugging interface, multiple LEDs and push buttons, as well as a selection of power supply connections.

Furthermore, the form-factor of the SENTIO-em core module has been chosen with simple deployability in mind. The module has a physical size of 75 mm × 49 mm, which allows for a screwless mounting of the platform in a Hammond 1591T plastic enclosure, as shown in Figure 2b. In the same way a sensor extension module (also shown in the picture) and a power supply unit can be stacked within the same box in a matter of seconds.

6) Interfaces: The SENTIO-em architecture includes several input and output interfaces. As previously mentioned, the core module provides interfaces for sensor extension, power supply connection, docking, as well as radio module integration. Furthermore, the module contains an SD-card slot that allows for local storage of large amounts of data in a size and cost-efficient manner. The user interface of the platform is rather simplistic, consisting of a reset button, a single programmable push button and a few status LEDs. In addition, the platform has been equipped with two magnetic switches, which allow user interaction with the sensor node without opening its physical enclosure. This, for example, allows to set the packaged system to a shipping mode, in which the node conserves energy until it has finally been deployed. For the usage of additional user interfaces, the platform can be placed on its laboratory docking solution.

C. SENTIO-em Software

For the development of the applications, a software stack, as depicted in Figure 3, has been implemented. For this, standard operating systems, such as TinyOS or Contiki, have been avoided, but the development of application code mainly relies on a C++ API. This allows for a uniform code, wherein individual hardware modules are interfaced via class instances, while keeping the overhead of the software stack at a minimum.

In order to support the development of larger applications, a state machine interface (SMI) has been developed on top of the platform's API. The SMI allows for the definition of application states in an abstract manner, as well as the state transition conditions. The state transitions can occur directly at the end of the current state (e.g., based on some condition), but are typically triggered by some external event, such as a timer, sensor or user interrupt. Furthermore, the state machine interface allows for the implementation of hierarchical finite state machines, which increases code readability and reusability in more complex applications. For more detailed information on the fundamentals of the SMI framework, we refer to [13].

IV. EXPERIMENTAL SETUP

For the evaluation of the SENTIO-em platform several tests have been conducted, which can be grouped into the following categories.

1) Platform evaluation under laboratory conditions
2) Platform evaluation with environmental influences
3) Platform comparison with existing solutions

While some previously presented system properties can only be described qualitatively (e.g., easy deployability), because of objectivity reasons, the presentation of results mainly focuses on quantitative data. Design decisions in compliance with qualitative constraints have been described in the previous section.

The primary quantitative parameter of interest is energy efficiency, which includes energy consumption, timings and computational performance. Additionally, the communication range is of interest in many outdoor sensor network deployments. However, as the communication range is dependent on the operational environment, and hence objective measurements are difficult to obtain, the radio module output power is analyzed instead, which allows for the estimation of the range in different environmental settings.

For energy consumption measurements, a fully implemented SENTIO-em platform (equipped with an SX1233 radio) is supplied from a constant voltage source and is configured for the respective system state under test. The current consumption of the platform is measured in the

 ISBN: 978-1-61208-296-7

supply line by means of an Agilent 34410A digital multimeter. The consumption measurements are conducted under laboratory conditions to estimate ideal behavior, as well as under different environmental influences. For the tests under environmental influence, a TestEquity 1000H environmental test chamber has been used in order to generate defined environmental conditions.

In a similar setup, platform timings have been evaluated. These measurements include the time for the system-wakeup, as well as transition times between different operating states. A TechTools logic analyzer has been used in order to conduct the measurements of the time periods. This logic analyzer provides a resolution of 10 ns.

In order to compare the energy efficiency of the SENTIO-em platform with that of existing platforms, we have analyzed the maximum clock frequency and the sleep-mode power consumption of a set of popular sensor nodes. While for SENTIO-em its obtained measurement results have been used, the data on the other nodes is extracted from the respective datasheets or research publications.

Finally, the RF output power of the Semtech SX1233 module implementation has been measured using an Anritsu MA24106A power meter.

V. Results

The evaluation results are presented separately in respect to energy efficiency and communication performance. All the presented results on the SENTIO-em platform are based on measurements.

A. Energy Efficiency

Energy efficiency in low duty-cycle applications, such as those that typically occur in the environmental monitoring domain, is determined by two platform parameters. These parameters are the power consumption of the sensor node during inactive periods and the time period the node has to remain active. In order to reduce the active time, in turn, the state transitions and the processing tasks have to be performed as quickly as possible.

Because the SENTIO-em platform is always operated at a constant supply voltage, the measurement of power consumption is performed by measuring the current consumption. The operating voltage of the platform is 3 V, which can be used to compute the power consumption, if this is necessary.

Table IV shows the results of the current consumption measurement for different operating states under laboratory conditions. All values represent the complete platform consumption and are measured as described in Section IV. During the measurements of the microcontroller consumption, the radio module is set to sleep mode, while the microcontroller is asleep (EM3) for radio measurements.

Overall the measured values comply with the expectations from the theoretical datasheet numbers. For example, in energy mode 3 (EM3) a current draw of 1.2 µA was measured, which is in accordance with the expected value (600 nA MCU + 500 nA voltage regulation + 100 nA radio transceiver). Special attention should be paid to the operating state *EM3 + RTC*, which represents the microcontroller in energy mode 3 operating with an external RTC. This state is the one most often used during sensor node inactive periods (i.e., between sensor samplings and radio activities). Due to the typically low duty-cycle in environmental monitoring applications, this value determines the lifetime of the sensor node, and thus closely correlates to that of the system lifetime. During transmission, the node's power consumption depends tremendously on the RF output power to be used, which, in turn, is defined by the distance between the transmitter and the receiver. The measured relationship between these two parameters is depicted in Figure 4.

Table IV
Energy consumption of SENTIO-em in its different operation states (Laboratory conditions)

State	Current	Condition
Active	6.1 mA	14 MHz
	8 mA	32 MHz
Low-Power	2.3 µA	EM1
	1.6 µA	EM2
	1.2 µA	EM3
	1.8 µA	EM3 + RTC
Radio	11–50 mA	TX
	16.5 mA	RX

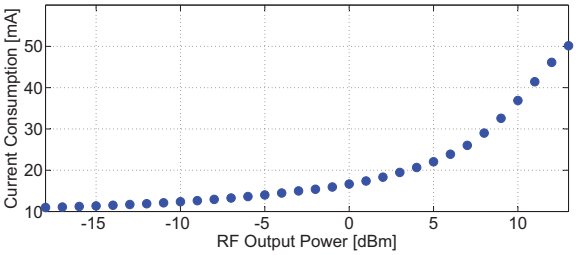


Figure 4. Current consumption of the Semtech SX1233 radio transceiver operating at different output power levels. Output power can be configured from −18 dBm to 13 dBm in 1 dBm steps

Because the majority of environmental monitoring applications take place in outdoor environments, the sensor nodes will be exposed to different environmental conditions. Therefore, conducting evaluations purely under laboratory conditions is insufficient. Figures 5 and 6 depict the results of measurements conducted under different temperature conditions. Figure 5 shows the platform's leakage current over a wide temperature range, whereas Figure 6 displays the temperature influence on active communication. All measurements have been conducted using a constant humidity setting of 40 %. In addition, the influence of different relative humidity conditions has been tested, but is ignored in this paper as its impact was found to be minimal. The

 ISBN: 978-1-61208-296-7

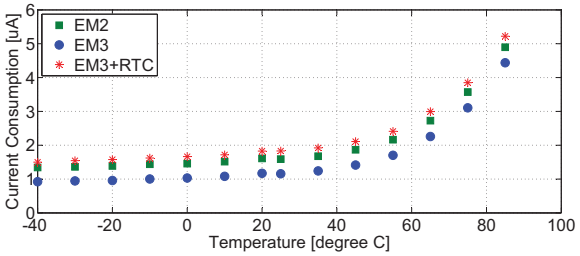


Figure 5. Current consumption of the SENTIO-em platform in its low-power modes over its entire temperature range. Humidity levels have been constant at 40 %

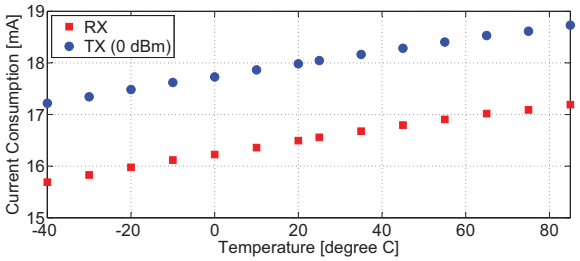


Figure 6. RF transceiver current draw during transmission (0 dBm output power) and reception under varying temperature conditions. Humidity levels have been constant at 40 %

measurement results show the expected increase of current draw with increasing temperature. While the temperature influence in the active communication shows a linear effect, in low-power modes the typical temperature dependency of leakage currents at a pn-junction can be observed. The leakage current in the platform remains low over the entire temperature range, but shows a drastic increase in relation to the laboratory conditions at temperatures above 40 °C.

Table V shows the transition times between the different operating states of the SENTIO-em platform. As many applications within the environmental monitoring domain follow a *sample-and-send* scenario with low duty-cycles (i.e., nodes periodically read sensor values and transmit them, while staying in low power modes as much as possible), transition times can play a significant role for the overall power consumption. The measurement results show that SENTIO-em can wake-up from low-power modes in less than 6 µs and be ready to transmit or receive in a few hundred µs. In combination with the typically low amounts of sensor data to be transmitted, these short transition times allow SENTIO-em to operate on very low duty-cycles.

Figure 7 indicates the energy efficiency of the presented platform in respect to popular existing sensor node platforms. In this figure, the sleep-mode power consumption of each platform is plotted against the maximum clock frequency, at which the respective platform can operate. The resulting location on the grid indicates how fast the platform can accomplish its tasks, and at which cost the processing performance comes.

Table V
TRANSITION TIMES BETWEEN DIFFERENT OPERATION STATES OF THE SENTIO-EM PLATFORM

From	To	Transition time
Sleep	Active	1.8 – 5.8 µs
Radio sleep	TX	150 µs
Radio sleep	RX	380 µs
TX	RX	390 µs
RX	TX	70 µs

B. Radio Performance

As mentioned previously, a parameter of interest concerning the radio transceiver is its communication range. This is particularly true in environmental monitoring applications, as the sensor networks might cover large outdoor areas. If the network does not have to simultaneously offer high spatial resolution, then long communication range can limit the system cost, as nodes with a purely relaying functionality can be omitted.

However, communication range is not a value that can be objectively measured. While it mainly depends on two transceiver parameters, the output power of the transmitter and the sensitivity of the receiver, several system external parameters will also influence the communication range. These parameters include the physical environment (i.e., objects in or close to the communication path, which result in multipath propagation, signal scattering or increased attenuation), as well as interference sources. As the external parameters are not under the direct control of the network/system operator and will vary from one location to the other, the evaluation of radio performance has been limited to the evaluation of transceiver properties.

Figure 8 shows the relationship between configured and measured output power of the SX1233 transceiver implementation, which indicates that no power is lost in the transmission line or the matching network.

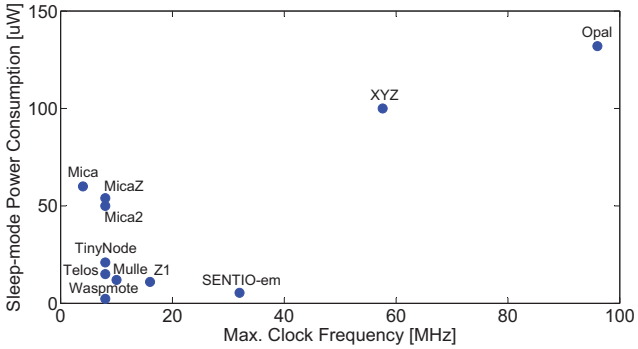


Figure 7. Comparison of different sensor node hardware platforms in respect to their energy efficiency. Energy efficiency is here measured by max. clock frequency and sleep-mode power consumption.

 ISBN: 978-1-61208-296-7

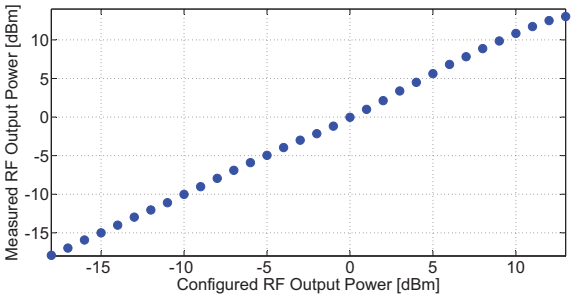


Figure 8. Comparison of theoretical and measured RF output power as an evaluation for radio transceiver implementation

VI. Conclusions

With SENTIO-em, we have presented a wireless sensor node platform particularly designed for the environmental monitoring application domain. While typical application-specific solutions limit the reusability of the node platform, which is hindering the research process, general purpose designs often lack the performance for the specific applications at hand. The application-domain-specific design of SENTIO-em, combines the advantages of both fields, which results in a platform that is optimized for a specific application domain, but which can be reused in different applications within the domain.

We have described the system architecture of SENTIO-em, which has been based on the typical application requirements of the environmental monitoring domain. These include long operating times, outdoor deployment, coverage of large areas, and low sampling rates. For the node platform, these requirements have been translated into system demands, such as energy efficiency, low quiescent current, long communication range, low hardware overhead, and simple deployability.

The implemented platform has been evaluated under both laboratory and different environmental conditions. The results show a reliable operation even under extreme temperatures, but also the typical temperature-dependent current draw of the involved semiconductor components, which will influence the lifetime of the system in different environments.

In comparison with typical existing platform solutions, SENTIO-em shows a high energy efficiency. The combination of fast processing times due to its 32 MHz clock frequency, short state transition times, and low power consumption in inactive states, proves the high energy efficiency of SENTIO-em, which determines the long system lifetimes desired in low duty-cycle applications within the environmental monitoring domain.

References

[1] J. Hill and D. Culler, “Mica: A Wireless Platform for Deeply Embedded Networks,” *IEEE Micro*, vol. 22, no. 6, pp. 12–24, Nov. 2002.

[2] J. Polastre, R. Szewczyk, and D. Culler, “Telos: Enabling Ultra-low Power Wireless Research,” in *Proceedings of the Fourth International Symposium on Information Processing in Sensor Networks*, 2004, pp. 364–369.

[3] H. Dubois-Ferriere, R. Meier, L. Fabre, and P. Metrailler, “TinyNode: A Comprehensive Platform For Wireless Sensor Network Applications,” in *Proceedings of the 5th International Conference on Information Processing in Sensor Networks*, 2006, pp. 358–365.

[4] P. Sikka, P. Corke, L. Overs, P. Valencia, and T. Wark, “Fleck - A Platform for Real-World Outdoor Sensor Networks,” in *Proceedings of the 3rd International Conference on Intelligent Sensors, Sensor Networks and Information Processing*, 2007, pp. 709–714.

[5] A. Mainwaring, D. Culler, J. Polastre, R. Szewczyk, and J. Anderson, “Wireless Sensor Networks for Habitat Monitoring,” in *Proceedings of the 1st ACM International Eorkshop on Wireless Sensor Networks and Applications*, 2002, pp. 88–97.

[6] K. Martinez, R. Ong, and J. Hart, “Glacsweb: A Sensor Network for Hostile Environments,” in *Proceedings of the First Annual IEEE Communications Society Conference on Sensor and Ad Hoc Communications and Networks*, 2004, pp. 81–87.

[7] G. Werner-Allen, J. Johnson, M. Ruiz, J. Lees, and M. Welsh, “Monitoring Volcanic Eruptions with a Wireless Sensor Network,” in *Proceeedings of the Second European Workshop on Wireless Sensor Networks*, 2005, pp. 108–120.

[8] M. Healy, T. Newe, and E. Lewis, “Wireless Sensor Node Hardware: A Review,” in *IEEE Sensors Journal*, Oct. 2008, pp. 621–624.

[9] P. Dutta, J. Taneja, J. Jeong, X. Jiang, and D. Culler, “A Building Block Approach to Sensornet Systems,” in *Proceedings of the 6th ACM Conference on Embedded Network Sensor Systems*, 2008, pp. 267–280.

[10] J. Artiola, I. Pepper, and M. Brusseau, *Environmental Monitoring and Characterization*. Elsevier Academic Press, 2004.

[11] K. Martinez, P. Padhy, A. Elsaify, G. Zou, A. Riddoch, J. Hart, and H. Ong, “Deploying a Sensor Network in an Extreme Environment,” in *Proceedings of the IEEE International Conference on Sensor Networks, Ubiquitous, and Trustworthy Computing*, 2006, pp. 186–193.

[12] T. Wark, W. Hu, P. Corke, J. Hodge, A. Keto, B. Mackey, G. Foley, P. Sikka, and M. Brunig, “Springbrook: Challenges in Developing a Long-term, Rainforest Wireless Sensor Network,” in *Proceedings of the International Conference on Intelligent Sensors, Sensor Networks and Information Processing*, 2008, pp. 599–604.

[13] M. Krämer, S. Bader, and B. Oelmann, “Implementing Wireless Sensor Network Applications using Hierarchical Finite State Machines,” in *Proceedings of the 10th IEEE International Conference on Networking, Sensing and Control*, 2013.

 ISBN: 978-1-61208-296-7

The Novel Microhotplate: A Design Featuring Ultra High Temperature, Ultra Low Thermal Stress, Low Power Consumption and Small Response Time

Hasan Göktaş, and Mona Zaghloul
Electrical and Computer Engineering
The George Washington University
Washington DC, USA
hgoktas.gwu@gmail.com, mona.zaghloul@gmail.com

***Abstract*— Microhotplate (MHP) has multiple applications, such as micro-heating elements, catalytic gas sensors, resistive gas sensing, infra-red source, non-dispersive infrared detector (NDIR) gas sensing, Fourier transform infrared (FTIR) spectroscopy, temperature sensing, flow sensing, pressure sensing, smart sensing and lab-on-chip. Low power consumption, high temperature, high thermal stability (low thermal stresses), uniform temperature distribution and fast thermal response are the most desirable features of the MHP design. This paper describes the design of the MHP that has ultra-high temperature (2000K), ultra-low thermal stress (90MPa), low power consumption (30mW) and small response time (9.13ms). All the results matched with Coventor simulation results.**

Keywords- Microhotplate; ultra high temperature; ultra-low thermal stress; low power consumption; small thermal response time.

I. INTRODUCTION

One of the popular usages for the MHPs is gas sensors. Fig. 1 shows the standard MHP gas sensor with the plate structure, a Polysilicon layer, between SiO_2.

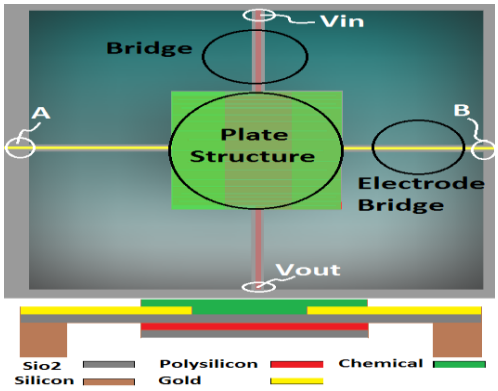


Figure 1. Microhotplate (MHP)

MHP is widely used for the catalytic and resistive gas sensing. According to Korotcenkov [1], high temperature would be desirable for selective gas sensors. It would reduce considerably the influence of air humidity and increase the sensitivity. Some metal oxide layers require high temperature for high sensitive gas sensors. For Ga2O3 sensor the operation temperature is 600-900 Celsius and it senses the O2 and CO gases. Oxygen sensors using SrTiO3 (Strontium titanate) operates at 1000 Celsius. According to [2], selectivity can be improved by operating the sensor in a temperature-modulated mode. High temperature MHP would allow a wide range of different gas sensing.

Nanomaterial processing is another one of the most important applications for high temperature MHP. It is especially important for CNT growth [3], thin film growth and characterization [4]. It requires high temperature operation with reconfigurable and controllable heating, and uniform surface deflection. All these features are well established in this paper.

Section-II concentrates on the causes of the thermal stresses and their possible solutions. Section-III demonstrates the effect of the spring structure and modulated AC signal on the low power consumption. Section-IV shows how to achieve the faster response time while keeping the uniform temperature distribution with uniform deflection. Section-V shows the novel MHP's performance by comparing it with the ones in the literature

II. LOW THERMAL STRESS MHP DESIGN

There are two causes of high thermal stress. The first is due to having different thermal expansion constant materials in the composite structure. The second is having a structure, which cannot expand when heated. Solutions in the literature are concentrated to solve the first cause by using either the same thermal expansion constant materials or the compatible materials in the composite structure. This solves the problem partially; however, fails at a high temperature around 700 Celsius. In this work, an analytical approach is first applied to the first cause to find the optimum dimensions solution (if any); and then, the novel structure is realized and applied to the second cause to get the optimum MHP for high temperature applications.

A. Thermal Stress Analysis

The equation for the composite structure having a different thermal expansion constant is given in [5] as shown below in (1)

$$\sigma_{xi}(y) = -\alpha_i E_i T_i(y) + E_i \frac{P_T I_{E2} - M_T I_{E1}}{I_{E0} I_{E2} - I_{E1}^2} + y E_i \frac{M_T I_{E0} - P_T I_{E1}}{I_{E0} I_{E2} - I_{E1}^2} \quad (1)$$

where

$$I_{E0} = \sum_{i=1}^{n} E_i w_i (y_i - y_{i-1}) \quad , \quad I_{E1} = \frac{1}{2} \sum_{i=1}^{n} E_i w_i (y_i^2 - y_{i-1}^2)$$

$$I_{E2} = \frac{1}{3} \sum_{i=1}^{n} E_i w_i (y_i^3 - y_{i-1}^3) \, , \, P_T = \sum_{i}^{n} \int_{y_{i-1}}^{y_i} \alpha_i E_i T_i(y) w_i dy$$

 ISBN: 978-1-61208-296-7

$$M_T = \sum_i^n \int_{y_{i-1}}^{y_i} \alpha_i E_i T_i(y) y \; w_i dy$$

The number of layers in a composite structure are represented in Fig. 2 and is denoted by i, where y_i-y_{i-1}=h_i represents the thickness of *i-th* beam, y_i represents the lower surface of *i-th* beam, E_i is elastic modulus, α_i is thermal expansion constant for *i-th* beam, and $\sigma_{xi}(y)$ is the stress value for *i-th* beam at thickness *y* and *w* represents the width.

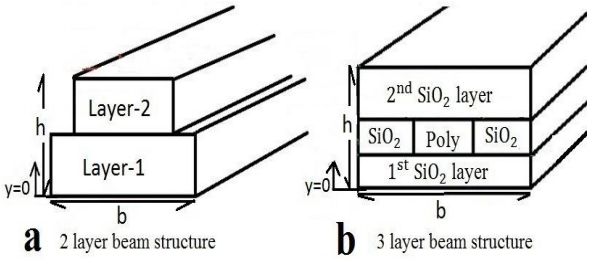


Figure 2. Cross section of the Composite structure

Conventional MHP structure is a composite structure. Fig. 3 shows the maximum thermal stress versus the temperature for the 3 layer composite structure (SiO2+Poly+SiO2). The stress value is around 240 MPa when the temperature reaches to 600 Kelvin. For the CMOS technology materials the maximum stress is around 100 MPa. Thus, at 600 Kelvin the CMOS MHP would be brittle with a conventional structure as shown in Fig. 1.

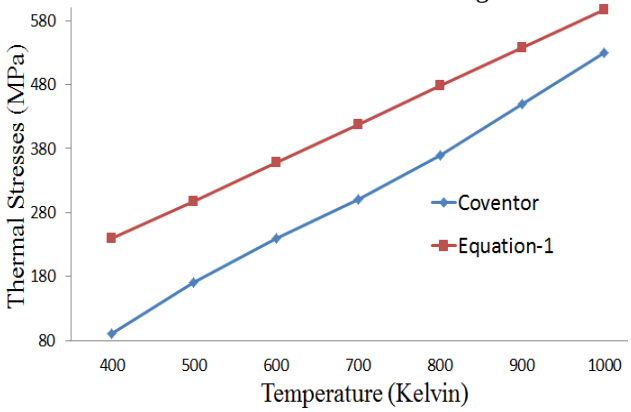


Figure 3. Stress versus Temperature for 3-layer MHP

Fig. 4a shows how the analytical results match with the Coventor simulation results for thermal stress in the middle of the polysilicon layer proving that the change in the width cannot solve the thermal stress problem. Fig. 4b shows that large thermal stress occurs in the center of the layer despite thermal stress decreasing on the sides proving changes in width cannot solve the thermal stress problem.

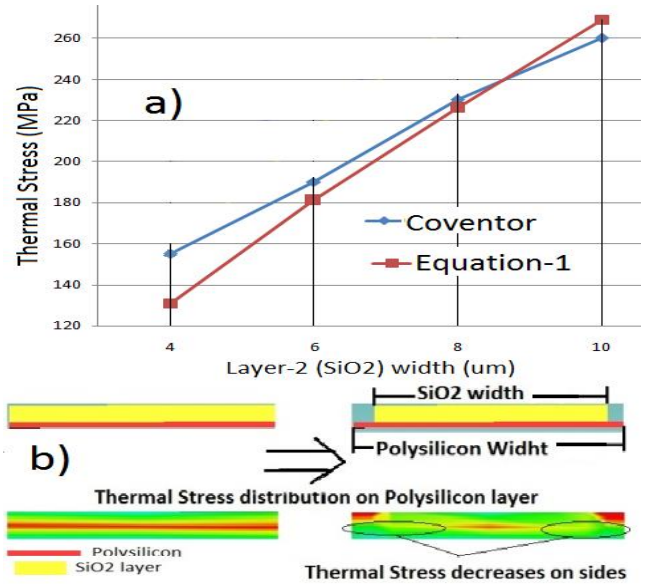


Figure 4. a) Effect of width on Thermal stress, b) cross-section view of the 2 layer composite structure

Fig. 5 shows a nice match between the calculation results and the Coventor simulation results for the thermal stress on the interface layer between SiO2 and Poly layer. In this result, layer 2 in Fig. 2a was used. In general, it applies to any number of layers in the interface. At first glance, it looks like the change in the thickness of one material in the composite structure can decrease the thermal stress.

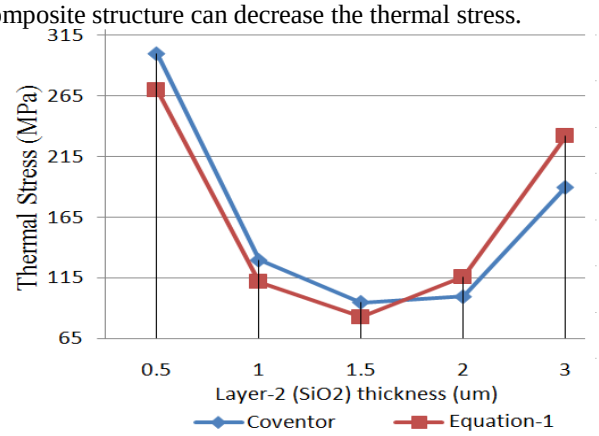


Figure 5. Effect of thickness on Thermal stress

Fig. 6 shows the distribution of the thermal stress throughout the thickness of several layers. Thermal stress is shown for the poly layer in Fig. 6a and for the SiO2 layer in Fig. 6c. Fig. 6b shows the (1)'s results. The match between the (1) and Coventor simulation results can be seen from Fig. 6. The graph in Fig. 6b shows the thermal stress versus thickness for SiO2 layer (y=0um to y=1.5um) and for the poly layer (y=1.5um to y=2um). It is proved by this example that changing the geometry (looking for the optimum dimensions) of the composite structure having a different thermal expansion constant is not a solution for low thermal stress. That is why a single layer structure or the composite structure having the same thermal expansion constant

 ISBN: 978-1-61208-296-7

materials would be a better option to decrease the thermal stress; however, it requires further optimzations as discussed in the next section.

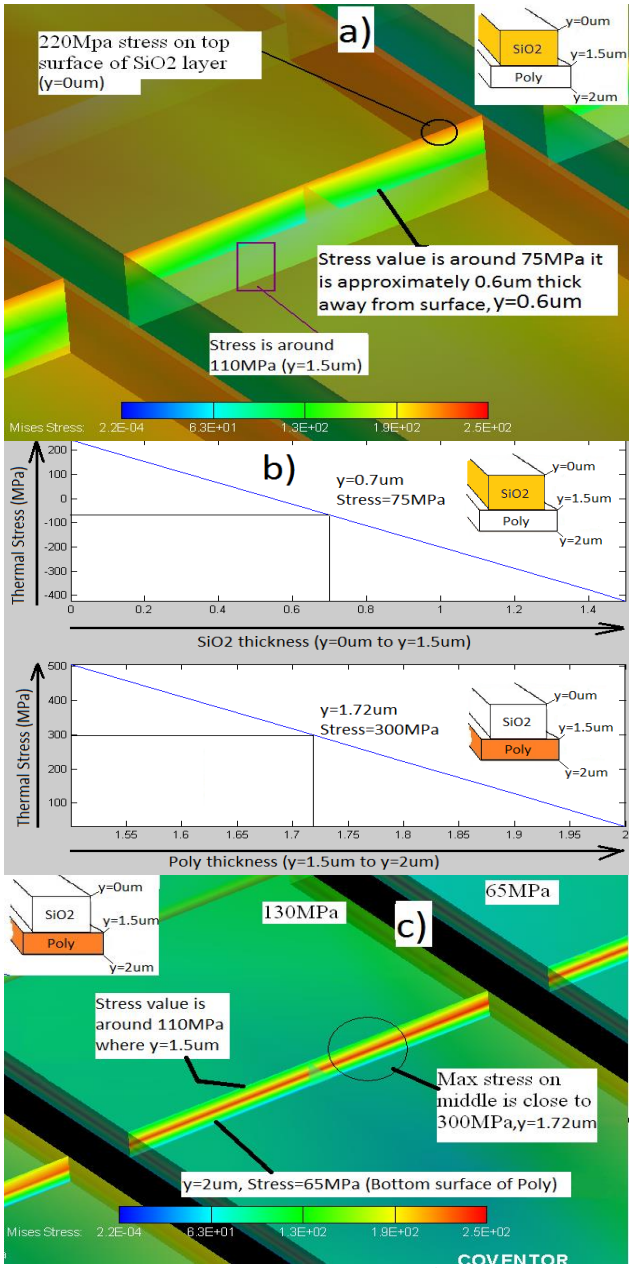


Figure 6. a)Thermal stresses versus thickness of SiO2, b)Thermal stress versus thickness from (1), c)Thermal stress versus thickness of Poly layer

B. *Designing a structure that can expand*

From the above discussion, a second cause for thermal stress is the restriction of expansion in the original MHP design shown in Fig. 1. A single layer and the composite structure with the same thermal expansion constant are described here. Spring structure allows the MHP to expand freely, and decreases the stress values, as shown in Fig. 7. Stress on the electrode bridge is around 370 MPa in Fig. 7a; however, it is around 230 MPa in Fig. 7b. Further design optimizations on the spring structure can be accomplished by varying the width, length, or thickness. If the width increases, then stress decreases according to equation $\sigma_{allowable} = \frac{M_{max}c}{(1/12)bh^3}$ and it increases the safety factor. In our design the maximum stress at 2000 Kelvin is less than 90 MPa as shown in Fig. 8. Fig. 11 shows the temperature effect on the thermal stress. The thermal stress increases while the thickness increases due to the increase in volume and thermal expansion. The safety factor is realized around 1.5.

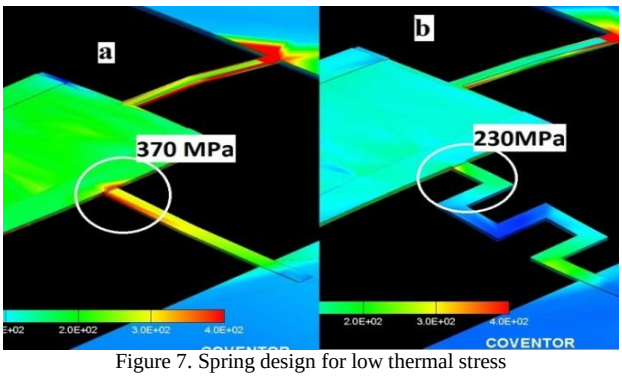


Figure 7. Spring design for low thermal stress

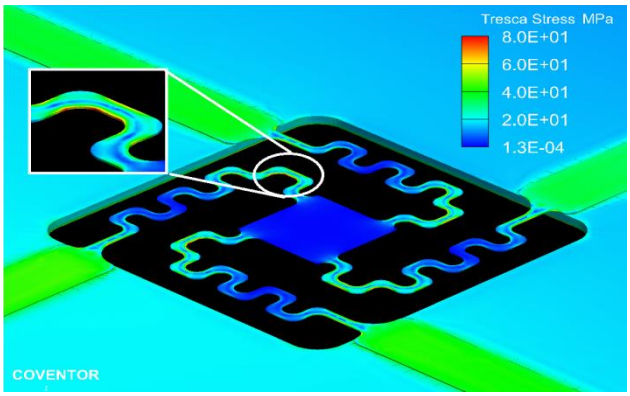


Figure 8. Maximum Thermal Stresses at 2020K

III. LOW POWER CONSUMPTION MHP

The spring structure decreases the heat flux to the substrate and keeps most of the heat on the MHP as shown in Fig. 9. R_o is the resistance of the heater plate, *R1* is for the heater bridge and *R2* is for the connection between the substrate and the heater bridge. T_o is the center temperature on the plate, *T1* is at the connection of the bridge and plate, *T2* is at the connection of the heater and substrate and *T3* is the substrate temperature. The temperature at each node can be found via thermal conduction and heat generation equations as shown below.

$$T_a - T_b = \frac{q'L^2}{2k}, \text{ where } q' = \frac{E_g'}{V},\ E_g' = I_{eq}{}^2 R,\ R = \rho\frac{L}{A} \quad (2)$$

$$T_a - T_b = q''R'', \text{ where } R'' = L/k \quad (3)$$

where q' is the heat generation per volume, V is the volume, R is the resistance, I_{eq} is the current flow, ρ is the

density, L is the length, A is the area, $q''$is the heat flux, k is the thermal conduction constant and R'' is the resistance for a unit surface area.

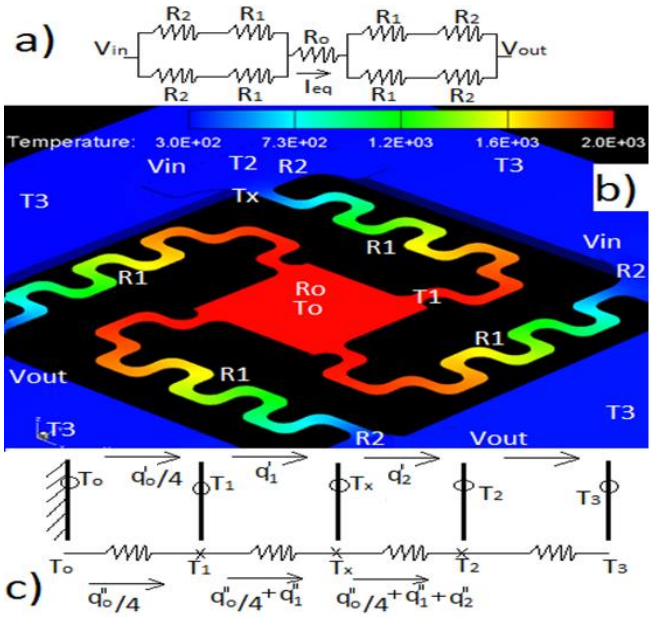


Figure 9. a) Electrical equivalent circuit of MHP, b) Temperature Distribution on MHP at 0.4Volt, c) Thermal Conduction equivalent circuit of MHP

Coventor simulation results matched with (2) and (3). Equation-2 is used in the section between T_o and T2. Equation-3 is used in the section between T2 and T3 as shown in Fig. 9c. This procedure allows us to find the optimum design in terms having low power consumption, high uniform temperature distribution and uniform deflection on the surface. The design in this paper has 37mW power consumption with a 0.46 Volt voltage at 2000 Kelvin.

Using AC voltage (frequency modulated signal) rather than DC voltage decreases the power consumption further for low-voltage design. Ta in Fig. 10 is the time period when the voltage is high; Tb is the time period when there is no voltage applied to the MHP. Tb depends on the cooling time of the system. In this way the system can be kept at 2000K with much more less power consumption.

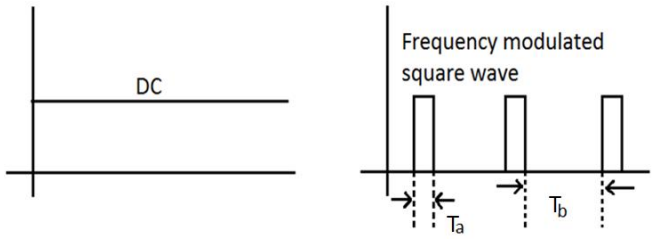


Figure 10. DC voltage versus modulated square wave

Fig. 11 shows an example for thickness effect on power consumption. A decrease in thickness also results in a decrease in power consumption because the resistance increases and drops the current flow.

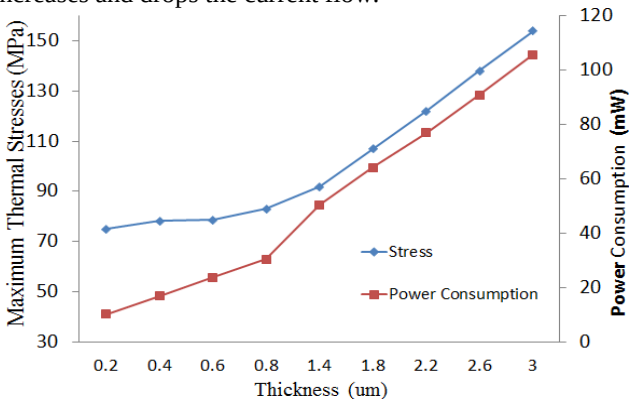


Figure 11. Power Consumption and Thermal stress versus thickness at 2000K

IV. SOLUTION TO FAST RESPONSE TIME WITH UNIFORM TEMPERATURE DISTRIBUTION AND DEFLECTION

The response time can be decreased by increasing the current because heat is proportional to the rate of current flowing through the heater. If the voltage increases more than the required value to reach the target temperature, then the heating rate will increase. This results in high temperature in a short time due to an increase in the current density. Fig. 12 shows how the voltage affects the temperature increase in a shorter time period. By combining the frequency modulated square wave solution from Fig. 10 with the high voltage solution from Fig. 12, we can get a reconfigurable heating rate on MHP. This feature is very important especially for high quality CNT growth. Having a single layer structure not only enables the uniform temperature distribution on the MHP as shown in Fig. 8b, but also enables a very small response time as 9.1ms to reach 2000K with 0.46V.

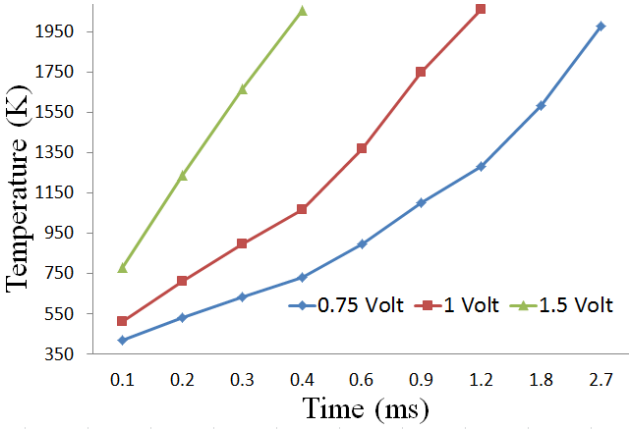


Figure 12. Transient Responses Temperature from Coventor simulation results when different voltages applied to the MHP

 ISBN: 978-1-61208-296-7

V. CONCLUSON

Previous results in the literature to the best of our knowledge show the smallest thermal stress is 210MPa at 650 Celsius in [6], the highest temperature is up to 1000 Celsius in [7], the lowest power consumption is 16mW at 550 Celsius in [8], and the smallest response time is 5ms for 400 Celsius in [9] and 10ms for 550 Celsius in [8]. We succeeded to design the MHP with less than 90MPa stress at 2000 Kelvin with 9.13ms response time, 37mW power consumption and 0.7um uniform deflection. Detailed results can be seen from Fig. 13. The MHP structure shown in this paper is currently being fabricated for measurements. MHP will be tested to sense the gases that require high temperature. The temperature modulated mode will also be tested by using the advantage of a wide range of temperature (0-2000K) capability to sense the different gases

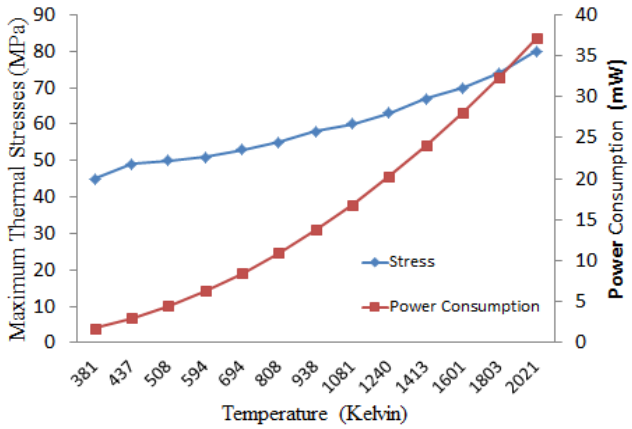


Figure 13. Power Consumption and Thermal stress versus temperature

REFERENCES

[1] G. Korotcenkov, "Metal Oxide for solid-state gas sensors: What determines our choice?" Material science and Engineering B 139 (2007)

[2] A. Vergara, E. Llobet, J. Brezmes, P. Ivanov, C. Cané, I. Gràcia, X. Vilanova, and X. Correig, "Quantitative gas mixture analysis using temperature-modulated micro-hot plate gas sensors: Selection and validation of the optimal modulating frequencies," Sensors and Actuators B 123(2007) 1002-1016

[3] M. S. Haque, K. B. K. Teo, N. L. Rupensinghe, S. Z. Ali, I. Haneef, S. I. Maeng, J. Park, F. Udrea, and W. I. Milne, "On-chip depositon of carbion nanotubes using CMOS microhotplates," Nanotechnology 19 (2008) 025607

[4] J. L. Hertz, D. L. Lahr, and S. Semanick "Combinatorial Characterization of Chemiresistive Films Using Microhotplates," IEEE Sensors Journal, Vol. 12, No 5, May 2012

[5] N. Noda, R. B. Hetnarski and Y. Tanigawa, Thermal Stresses, Second Edition, Tylor & Francis, 2003, pp. 39-54

[6] J Puigcorb´e, D Vogel, B Michel, A Vil`a, I Gr`acia, C Can´e and J R Morante, "Thermal and mechanical analysis of micromachined gas sensors," J. Micromech. Microeng. **13** (2003) 548–556

[7] L. Mele, F. Santagata, E. Iervolino, M. Mihailovic, T. Rossi, A.T. Tran, H. Schellevis, J.F.reemer and P.M. Sarro, "Sputtered Molybdenum as conductive material for high-temperature microhotplates," Transducers'11, Beijing, China, June 5-9, 2011

[8] P.K. Guha, S.Z. Ali, C.C.C. Lee, F. Udrea, W.I. Milne, T. Iwaki, J.A. Covington and J.W. Gardner, "Novel design and characterization of SOI CMOS micro-hotplates for high temperature gas sensors," Sensors and Actuators B 127 (2007) 260–266

[9] Lei Xu, Tie Li, Xiuli Gao, and Yuelin Wang, Senior Member, IEEE, "Development of a Reliable Micro-Hotplate With Low Power Consumption," IEEE Sensors Journal, vol. 11, no. 4, April 2011

 ISBN: 978-1-61208-296-7

A Miniaturized 4-Channel, 2KSa/sec Biosignal Data Recorder With 3-Axis Accelerometer and Infra-red Timestamp Function

Jim Austin, Chris Bailey,
Anthony Moulds,
Garry Hollier, Michael Freeman
Department of Computer Science
University of York, UK
jim.austin@york.ac.uk
chrisb@cs.york.ac.uk
anthony.moulds@york.ac.uk
garry.hollier@york.ac.uk
mike.freeman@york.ac.uk

Jim Austin, Alex Fargus,
Thomas Lampert
Cybula Ltd,
York, UK
austin@cybula.com
alex.fargus@york.ac.uk
tom.lampert@hotmail.com

Bettina Platt, Gernot Riedel
University of Aberdeen,
Institute of Medical Sciences
Aberdeen, UK
b.platt@abdn.ac.uk
g.riedel@abdn.ac.uk

***Abstract*—In this paper, we describe a miniature biosignal data sensor and recorder device, (NAT-1) with 3-axis accelerometer, and a 2KSa/sec all-channel recording capacity of 24 hours or more with a single zinc-air battery cell. The device measures less than 18x22mm and weighs less than 2.3 grams, including the battery. The device is evaluated in several contexts, including electroencephalography data collection, forearm electromyography capture, and in use as an accelerometer for movement capture in sport. We describe the device, its use, and show examples of some monitoring applications; in all cases, data is collected in an untethered stand-alone mode. The NAT-1 has already achieved translation to commercialization and is the first of a new family of sensors.**

Keywords-Neurophysiology; EEG Recorder; Bio-signal sensors; Medical sensors.

I. INTRODUCTION

The use of biosignal data acquisition is increasingly important in many applications spaces, not least of which is biomedical applications. The use of devices to record EEG (Electroencephalography), EMG (Electromyography), ECG (Electrocardiography) and movement is of particular interest. Often, these measurements are taken in wired or wireless umbilical modes, in other words, within a clinical evaluation setting, with data captured and analyzed over relatively short time windows, and in unnatural settings.

The ability to perform ambulatory monitoring of patients, on the other hand, provides the possibility of long-duration data capture of bio-parameters in a normal living-condition or work-place. This has been an aim for many decades, and it is interesting to contrast how this has been achieved in past and present technologies; ranging from early magnetic tape based data capture [1], digital systems [2], Custom Integrated Circuits [3,4] and medical data recorders [5,6]. Such capabilities are identified by many clinical researchers as being desirable, and the opportunity to learn more about medical conditions, as well as the condition of individual patients themselves is seen as a major motivator for developing suitable devices.

The motivation of our project "NAT" (Neural Activity Tracker) is to produce a multi-purpose data sensing and recording solution that is extremely small, lightweight and having a recording capacity of days to weeks, dependent upon the selection of parameters such as sample frequency. The NAT is such a device, and it is this which we describe in the remainder of this paper. We believe that the device, even in its current implementation, is capable of being used in many applications, including non-obtrusive use in wearable patient monitoring.

This paper presents the NAT in detail in Sections 2, 3 and 4, whilst Section 5 documents some initial experimental applications of the device. Section 6 highlights the additional resources developed alongside NAT, including docking station, control software and data analysis tools, and the infra-red data signature time-stamp daughterboard. A brief state-of-the-art context is given in Section 7, and conclusions in Section 8.

II. A SMALL FORM-FACTOR DATA-RECORDER

The Neural Activity Tracker (NAT) is a device only 18x22mm in size, less than 10mm in height (primarily due to the battery module), and weighs less than 2.3 grams, which means it is attractive in applications where regulatory constraints apply (e.g., use with small animals). It is so light in weight that it is almost unnoticeable in normal use as a human-subject wearable body-sensor. This low-weight attribute also means that multiple devices can be worn individually or in small groups where appropriate, without causing encumbrance of the subject's normal movements and behavior. The NAT is shown in the image of Fig. 1, with a ball-point pen of normal size for scale. In the photograph provided in Fig. 1, one can see some interesting features. First of all, the major part of the device has a profile of less than 4 mm. Also visible in Fig. 1 is a specially designed battery clip for housing a zinc-air cell (as used in hearing aids). This is the gold-plated metal structure. The use of flatter button cells can be envisaged to make a smaller profile possible, though this has weight implications.

 ISBN: 978-1-61208-296-7

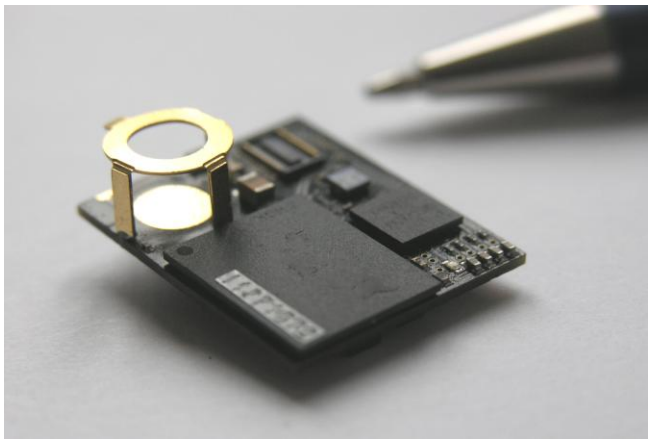


Figure 1. Top-view of NAT-1 device

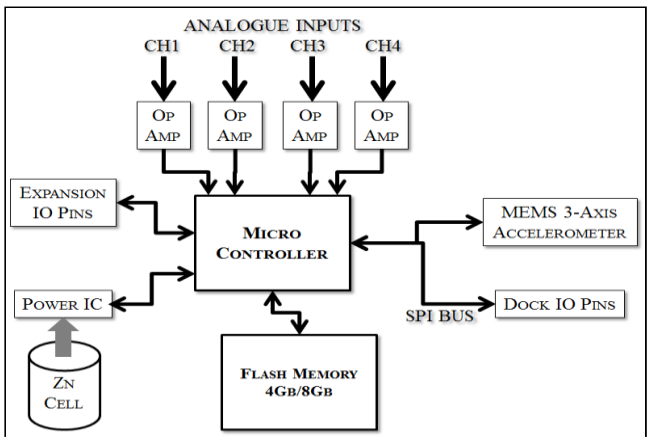


Figure 2. NAT Device – Sub-component Block Diagram.

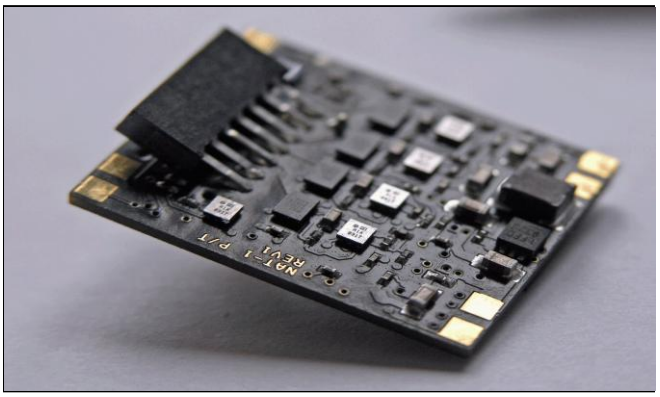


Figure 3a. Analogue Connector Port on NAT-1 Underside.

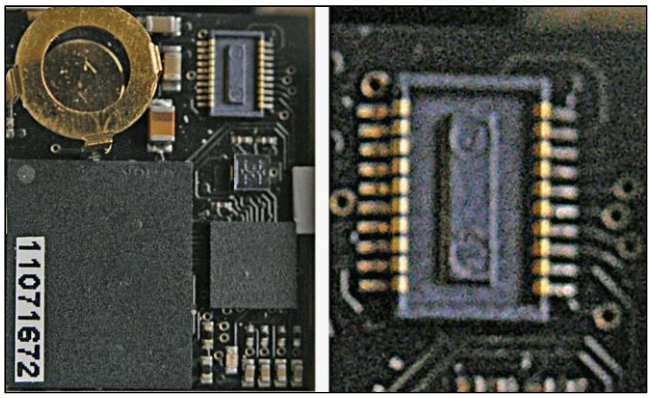


Figure 3b. Overhead view of NAT-1 and close-up of Daughterboard

Near the tip of the ball-point pen, one can see a small low-profile connector socket, which is actually a mezzanine daughterboard connector port. This permits a range of possible extension modules. Fig. 2 shows the system level block diagram.

The NAT device comprises of three key components – a flash memory chip (largest chip in Fig. 1), a proprietary CPU (mid-sized chip of Fig. 1), and a proprietary MEMS (MicroElectroMechanical System) device for accelerometer (smallest chip of Fig. 1). Additional analogue front-end components provide appropriate signal conditioning for the signal ranges typically encountered in a range of biosensor and biomedical applications.

III. ZINC-AIR POWER CELL DESIGN CONSIDERATIONS

A particular feature of NAT-1 is the use of a single zinc-air power cell. Such cells are widely used in hearing aids, and have an active power-delivery life-span of around two weeks, after which the cell begins to lose its effectiveness. This degradation begins as soon as the air-seal is removed from the cell, but allows for long-duration continuous power delivery from a very small and light power cell.

Because of the zinc-air reaction mode of power generation in the specific cell mentioned, it is not possible to draw power above a certain ceiling, which is limited by the capacity for air circulation into and out of the cell rather than its lifetime power delivery capacity. Consequently, it is necessary to spread power utilization to avoid frequent peak-power drainage. Failure to do so results in cell depletion and CPU brown-out. This is unlike a standard power-cell such as a button-cell, where current is limited by the device's ability to operate within a safe discharge rate rather than by a continuous power availability driven by an external reaction factor. Nonetheless, we have developed the NAT-1 with a continuous recording capacity in excess of 24 hours at full design-spec sample rates (see later power data for details).

IV. DEVICE SPECIFICATIONS

The NAT-1 has the specifications as outlined in Table 1. The device has a wide range of possible sample rates, ranging from 100Sa/sec to 2KSa/sec via the user interface software application. At 2KSa/sec, the device consumes 4.8mA of current from a single 1.4-volt cell, and can record for up to 12 hours. An 8-GBit flash option is possible and would have up to 24 Hours of recording capacity at maximum rates. At lower sample rates, the capacity of the

 ISBN: 978-1-61208-296-7

flash is extended to many days, and power consumption reduces to allow for longer recording times. This compares well with reported state of the art [3,4], given that the system is comprised of readily available commodity integrated circuits. The device has three important connection mechanisms, these being the analog input connector (angled connector block to left of device in Fig. 3a), the daughterboard extension socket (Fig.3b) and the docking-shoe connectors (seen in Fig. 3a as small gold contacts at the board corners).

V. DEVICE USEAGE EVALUATIONS

To date, most of the testing of the device has been limited to engineering development contexts, but using live EEG and EMG data collection with additional test scenarios. This does not include validation in a clinical certification context.

Three test scenarios are reported. EEG and Accelerometer data collection from mice in live test scenarios has been undertaken by researchers at the University of Aberdeen. This work continues and will build a validation case for clinical use. We have undertaken preliminary EMG testing scenarios at the University of York, mainly to evaluate the signal range suitability for gel-electrode signal capture.

TABLE I. NAT-1 SPECIFICATIONS

Parameter	Limits	Units
Analogue inputs	4	channels
Bits per channel recorded	11	bits
Accelerometer	3	Axis
Bits per Accel. Axis	8	bits
Sample rate (max)	2000	Sa/Sec x 4 ch
Max Current 2KSa/s 500Sa/s	4.8 2.4	mA mA
Data Capacity	4 or 8	Gbits
Recording Time at 2KSa/Sec	12 or 24	Hours
Analogue Range	± 1000 ± 4000	uV (NAT-1) uV (NAT-2)
Accelerometer range	Selectable 2 or 8	G (G-force)
Accelerometer sensitivity	18 72	mG at 2G range mG at 8G range

To test the accelerometer capabilities, the device has been used as a golf-club attachment to capture golf swing behaviors in terms of club rotation, side-movement and swing-path motion. In all cases, the device was used untethered and powered exclusively with a single zinc-air cell. These cases are shown in our test data of Figs. 4a, 4b, 4c and 4d, and are described in the following sub-sections.

A. EEG Measurements

Data has been captured with a wide range of sample rates, ranging from below 200Sa/sec to 2000Sa/sec. Whilst it is generally accepted that useful frequency ranges for EEG are in the region below a few hundred Hertz, the advantage of a higher frequency sampling regime comes from the ability to determine precise timing and time-separation of events, either within the EEG and Accelerometer data streams, or from an external trigger action. The NAT-1 has an extension board that can record Infra-Red-Timecodes within the data stream and this can be used to record trigger events and subsequent EEG responses with high accuracy in the time domain. Higher frequency sampling also allows for improved noise performance, since downsampling and digital filtering can be applied to the data stream after capture.

Experimentally, it has been found that the device performs well with mice as subjects, as expected. Rats however have a larger EEG signal range, and the NAT-1 device was originally not designed for this purpose. However, the NAT team is currently implementing a new programmable gain function to permit a wider range of signal input ranges to be captured without saturation or clipping artifacts. This NAT-2 device is undergoing testing at the time of writing and is not reported here. Examples of mouse EEG data captured with NAT-1 are given in Fig. 4a, which also shows accelerometer and IR-event code side channel. It can be seen that the IR event code correlates to regions of increased EEG activity. This NAT-evaluation data was provided by a well-established team with a long experience of using rodent EEG data recorders [7,8,9].

B. EMG Measurements

Initial experimentations with EMG measurement are promising. The device was attached to a subject's forearm and used to measure repeated muscle contraction-relaxation cycles and simultaneously record 3-axis accelerometer readings. The device was connected directly to standard clinical Ag/AgCl disposable adhesive surface electrodes. Again, the NAT-1 was not originally intended for this application, but we wished to evaluate its potential and identify improvements required. For this experiment, no pre-amplification or attenuation was applied to the electrode feeds, and as would be expected this means that the signal range captured does not perfectly match the analogue input range of the NAT-1 device in this initial setup. Nonetheless, it can be seen in Fig. 4a, that the muscle contraction results in repeatable EMG signals. It is also seen in Fig. 4a that clipping occurs frequently.

 ISBN: 978-1-61208-296-7

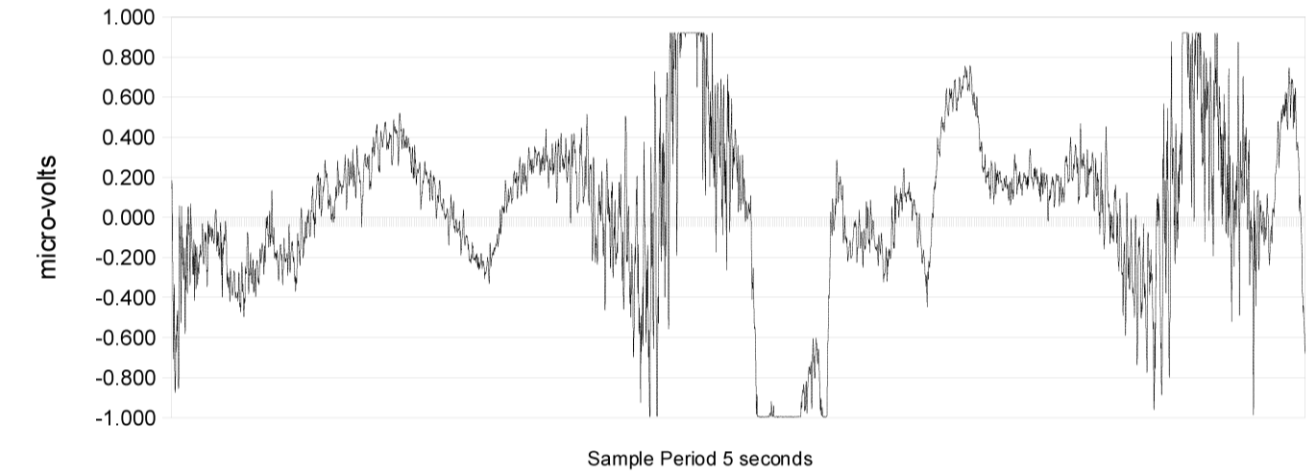


Figure-4a. Electromyography test case, repeated bicep contraction regime, with gel-electrode contact.

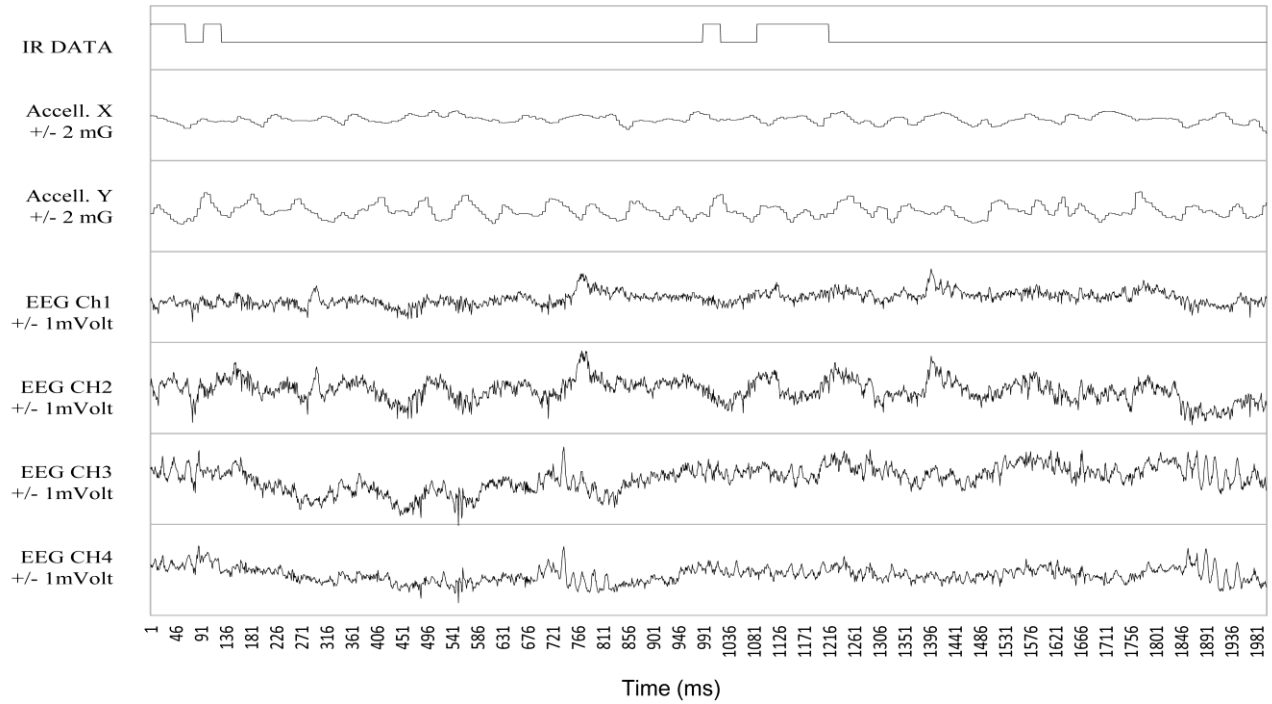


Figure 4b. EEG data collection in tandem with 2-Axis Accelerometer. And Infra-Red Pulse Code Recording.
(top trace – IR channel, mid - 2 axes of accelerometer , and bottom – 4 EEG channels)

In this experimental set-up with no attempt to pre-scale the input signal to the NAT-1 input range, it is not surprising to see this effect. The revised device, which incorporates programmable gain features, will allow for such signal measurements to be properly scaled and ranged to permit full-scale capture without clipping. The combination of accelerometer, and EMG recording in a very small lightweight package make this an attractive device for monitoring of symptoms such as essential tremor, Parkinson's disease, and other symptoms. That application will be a high priority for the York group's future focus.

C. Golf-Swing Accelerometry

This experiment was conducted in order to gather accelerometer data with a full-scale and saturation condition, including under extreme-G conditions. The device survived intact and retained battery supply integrity throughout the testing, due to the specially designed battery retaining clip. It might have been expected that the extreme forces of impact might lead to physical battery disconnects. However, no evidence was observed in the test data, indicating that the clip was highly effective.

 ISBN: 978-1-61208-296-7

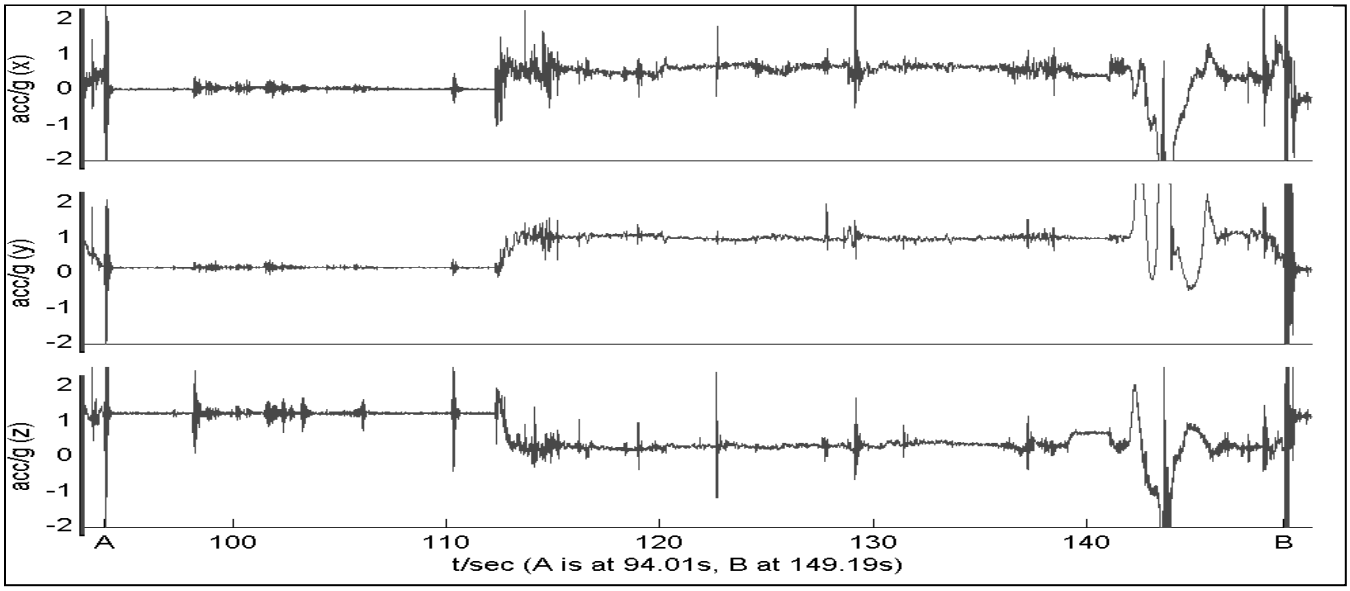


Figure 1. Top-view of NAT-1 device

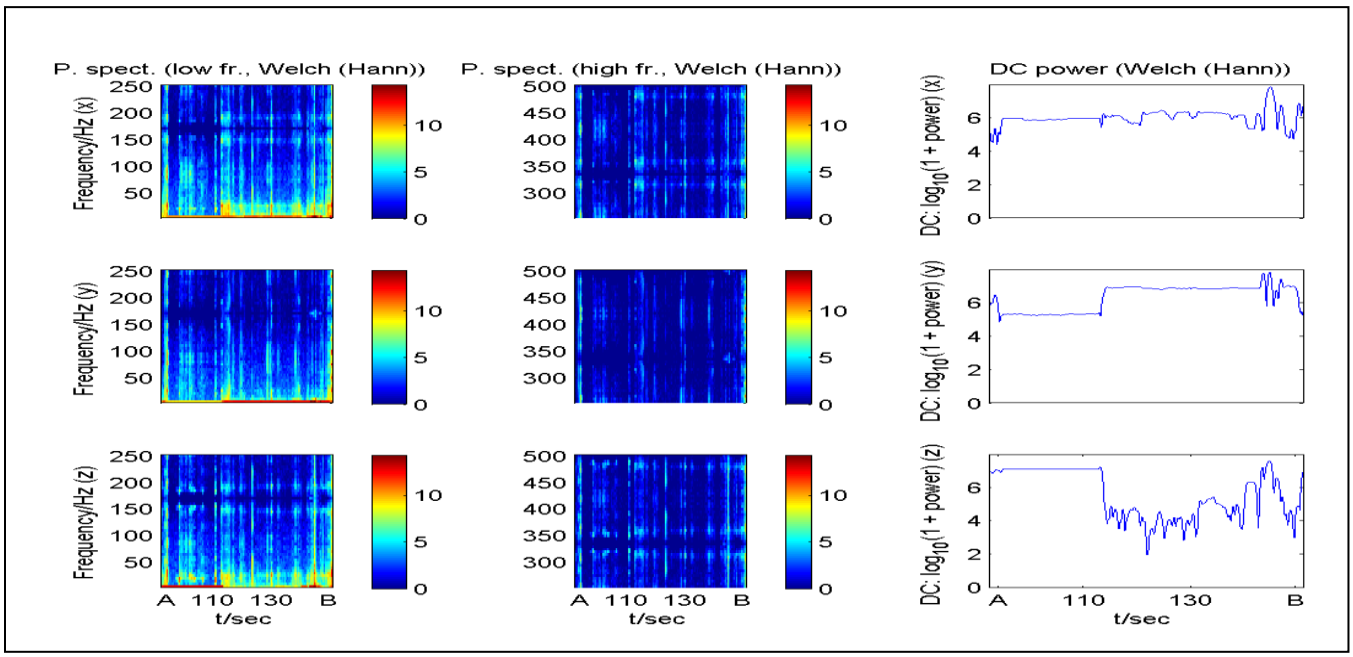


Figure 4c. Accelerometer data collection for Golf Club Swing Test (*Sample rates set to 1500 Hz, with (right) Spectral Plots using Welch function)*

The sensor was attached to a golf club at the mid-point of the club shaft (not at the head). This provides a good measurement point to permit shaft-rotation to be measured on one axis, the back/forward swing on a second axis, and any side motion to be captured by a third axis. The data was gathered from 100 golf-swings of varied competence. Examples are shown in Fig. 4c.

The golf data can be examined and can be seen to represent a back-swing and drive, followed by impact. This data was further analyzed using a spectral analysis method to yield the spectral time-series graphs shown in Fig. 4d. The data was processed using a Welch algorithm [10] and Hann windowing function [11].

VI. DOCKING STATION, ANALYSIS TOOLS, AND EXTENSION DAUGHTER BOARDS

The NAT device has the advantage of being provided with a bespoke docking station, for download of data from flash via USB. This is illustrated in Fig. 5. Once downloaded, the data can be processed using Cybula Ltd Signal Data Explorer software suite, which can be trained to perform auto detection and classification of signal behaviors and events (see screenshot in Fig. 6). The device programming tool is shown in Fig. 7, and shows a screenshot of the user interface allowing programmability of a wide range of device features.

 ISBN: 978-1-61208-296-7

A final aspect of the current NAT-1 system suite is the infra-red time-code recording daughterboard. This extremely small add-on (see Fig. 8 for a picture of devices with coin for scale) allows recording of an infrared pulse code stream along-side the analogue channels and accelerometer.

For NAT-1 devices, one accelerometer axis is sacrificed to permit IR time-stamping data recording. The pulse codes recorded are generated from a custom IR key-coding terminal or from triggered sensors in a maze, for example. These may be decoded later by software data analysis.

Future developments for the daughterboard connector could include RF telemetry functions. It is also possible to use the daughterboard connector to act as a micro-backplane to permit multiple NAT-1 devices to be ganged and synchronized to perform something like 16 or 32 channel acquisition, by linking multiple NATs in a synchronization.

VII. BRIEF CONTEXT AND STATE OF THE ART

The use of ambulatory EEG monitoring, be it in animal behavior studies or in human patients, is increasingly valuable. This has importance for the role of extended out-patient monitoring of EEG as a diagnostic aid and guide to treatment [12,13]. Recent work in this field includes many wireless approaches to ambulatory monitoring, but these are effectively 'tethered' via a secondary monitoring or recording station [14,15]. Recent literature reports an 8-channel EEG sensor node which “measures 35mm x 30mm x 5mm excluding Li-ion battery” [14], and power consumption in the range of 27-42mW, implying around 35mW for a four-channel system. The NAT, in contrast, consumes approximately 3.4mW with a 4-channel 500Sa/sec sampling regime. This is in part due to the elimination of wireless connection, though the use context of the two devices do not completely overlap – NAT is primarily designed for offline data analysis. Meanwhile, a full-ASIC dedicated solution can outperform NAT in terms of power consumption [15]; but, again, this generally requires a wirelessly tethered auxiliary system in the locale of the patient at all times to receive data. On-board classification and signal event detection has been used to reduce the RF overhead in some devices and systems, such as on-board seizure detection [4]. However, there is significant benefit in having a continuous recording of live data and especially where the purpose is to gather data to inform the development of such diagnostic/classification approaches.

VIII. CONCLUSIONS AND ONGOING WORK

We believe that the NAT-1 is a versatile and very compact device, with a low weight and long-duration high frequency recording capacity. Its application in areas such as EEG, EMG and in more complex monitoring scenarios, such as tremor-motion and EMG signaling in patients with conditions, such as Parkinson's disease, are the areas of greatest interest for this project.

Figure 5. NAT USB docking station

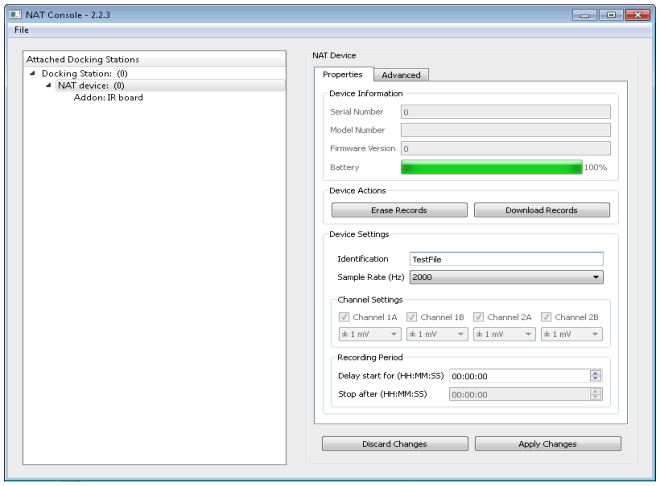


Figure 6. SDE (Signal Data explorer) Software toolset developed by York/Cybula

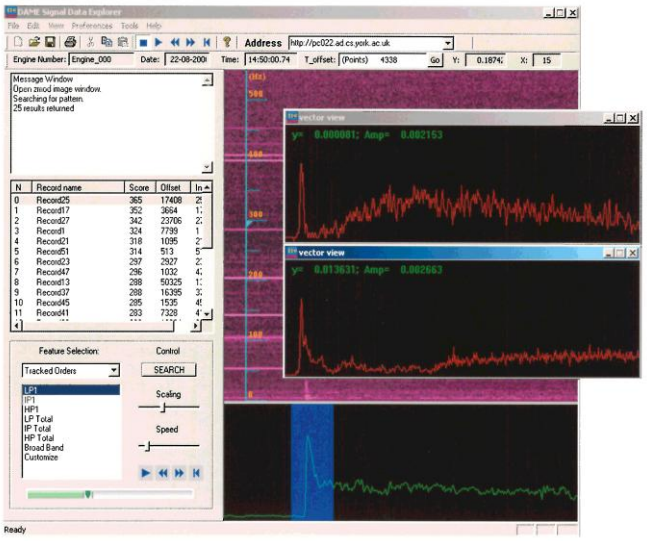


Figure 7. Control Panel Umbilical software showing typical device configuration parameter entry panel.

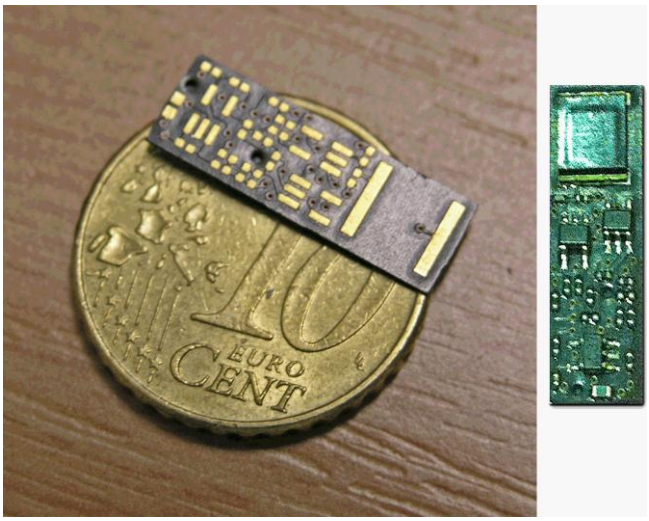


Figure 8. NAT Infra-red Receiver Daughter-Board

The York team [16] and partners CYBULA Ltd. [17] are very interested in developing collaborative partnerships and would welcome enquiries from prospective evaluators and end-users.

Current improvements are focusing upon a revised device design with individually programmable channel gain capabilities. This will allow a wider range of applications and signal ranges to be accepted by the device without the clipping observed in our preliminary EMG tests for example. It also allows for the possibility of multi-sensor recording, with for example scalp electrode EEG on several channels and EMG input with different gain settings on another channel.

The device has also been operated in a live-streaming mode, and to date this has allowed the development of a monitoring station for observing signal behaviors acquired by the device in situ.

Finally, the use of multiple devices is an area we have further interest in expanding. Ganging of multiple devices with synchronization via the expansion connector would allow arbitrary channel combinations, whilst in a live streaming mode, acquiring and logging data from multiple devices simultaneously is possible (though the device is ultimately intended as an untethered ambulatory system unit).

ACKNOWLEDGMENT

The NAT-1 device has been developed through a number of collaborative activities, including Cybula Ltd, University of York, and University of Aberdeen. We are also grateful for the support of the UK Government's Technology Strategy Board funding agency in supporting the development of the device.

REFERENCES

[1] J. R. Ives and J. F. Woods, "4-Channel 24 hour cassette recorder for long-term EEG monitoring of ambulatory patients," Journal of Electroencephalography and Clinical Neurophysiology 39.1, 1975, pp. 88-92.

[2] A. L. Vyssotski, A. N. Serkov, P. M. Itskov, G. Dell'Omo, A. V. Latanov, D. P. Wolfer, and H. P. Lipp, "Miniature neurologgers for flying pigeons: multichannel EEG and action and field potentials in combination with GPS recording," Journal of Neurophysiology, vol. 95, no. 2, 2006, pp. 1263-1273.

[3] R. F. Yazicioglu, P. Merken, R. Puers, and C. Van Hoof, "A 200μW eight-channel acquisition ASIC for ambulatory EEG systems," Solid-State Circuits Conference, 2008. ISSCC 2008, San Francisco. IEEE International Digest of Technical Papers, Feb 2008, pp. 164,603.

[4] N. Verma, A. Shoeb, J. V. Guttag, and A. P. Chandrakasan, "A Micro-power EEG acquisition SoC with integrated seizure detection processor for continuous patient monitoring,", Symposium on VLSI Circuits, Kyoto, Japan, 2009, pp. 62,63.

[5] E. Waterhouse, "New horizons in ambulatory electroencephalography," IEEE Engineering in Medicine and Biology Magazine, vol. 22, no. 3, 2003, pp. 74-80.

[6] J. Fu-Shan, Y. Tseng, and J. Jang, "Modular design of a long-term portable recorder for physiological signals," Measurement, vol. 43, no. 10, 2010, pp. 1363-1368.

[7] A. Jyoti, A. Plano, G. Riedel, and B. Platt, "EEG, Activity, and Sleep Architecture in a Transgenic AβPP swe/PSEN1 A246E Alzheimer's Disease Mouse," Journal of Alzheimer's Disease, vol. 22, no. 3, 2010, pp. 873-887.

[8] B. Platt, B. Drever, D. Koss, S. Stoppelkamp, A. Jyoti, A. Plano, A. Utan, G. Merrick, D. Ryan, V. Melis, H. Wan, M. Mingarelli, E. Porcu, L. Scrocchi, A. Welch, and G. Riedel, "B, PLB1," PloS one, vol. 6, no. 11, 2011, pp. 1-13.

[9] B. Platt, B. Drever, D. Koss, S. Stoppelkamp, A. Jyoti, A. Plano, A. Utan, G. Merrick, D. Ryan, V. Melis, H. Wan, M. Mingarelli, E. Porcu, L. Scrocchi, A. Welch, and G. Riedel, "Epidural EEG Recordings Using Microchips in Behavioral Context," Proceedings of Measuring Behavior 2010, Eindhoven, The Netherlands, August 2010, pp. 26-29.

[10] P. D. Welch, "The Use of Fast Fourier Transform for the Estimation of Power Spectra: A Method Based on Time Averaging Over Short, Modified Periodograms," IEEE Transactions on Audio Electroacoustics, (AU-15,1967), pp.70–73.

[11] F. J. Harris, "On the use of windows for harmonic analysis with the discrete Fourier transform," Proceedings of the IEEE , vol. 66, no. 1, January 1978, pp. 51,83.

[12] S. A. Rizvi, J. F. Téllez-Zenteno, S. L. Crawford, and A. Wu, "Outpatient ambulatory EEG as an option for epilepsy surgery evaluation instead of inpatient EEG telemetry," Epilepsy & Behavior Case Reports, vol. 1, 2013, pp. 39-41.

[13] H. J. Faulkner, H. Arima, and A Mohamed, "The utility of prolonged outpatient ambulatory EEG," Seizure, vol. 21, no. 7, 2012, pp. 491-495.

[14] S. Patki, B. Grundlehner, A. Verwegen, S. Mitra, J. Xu, A. Matsumoto, R. F. Yazicioglu, and J. Penders, "Wireless EEG system with real time impedance monitoring and active electrodes," IEEE Biomedical Circuits and Systems Conference (BioCAS), Hsinchu, Taiwan, 2012, pp. 108,111.

[15] S. Mitra, X. Jiawe, A. Matsumoto, K. A. A. Makinwa, C. Van Hoof, and F. Yazicioglu, "A 700μW 8-channel EEG/contact-impedance acquisition system for dry-electrodes," 2012 Symposium on VLSI Circuits (VLSIC), Honolulu, Hawaii, 2012, pp. 68-69

[16] Advanced Computer Architectures Group (ACAG). www.cs.york.ac.uk/arch , last accessed 07/29//2013.

[17] Cybula Limited. UK Limited Company. www.cybula.com , last accessed 07/29/2013.

 ISBN: 978-1-61208-296-7

Multi Sensor Atmospheric Icing Station Performance in Cold Climate- A Case Study

[1] Muhammad S. Virk, [2]Taimur Rashid, [3]Umair N. Mughal, [4]Kamran Zaman. [5]Mohamed Y. Mustafa

Department of Technology, Narvik University College, Narvik, Norway
Email: [1]msv@hin.no, [2]timavion58@gmail.com, [3]unm@hin.no, [4]kamimehr@yahoo.com, [5]myfm@hin.no.

***Abstract*— To amicably acquire the data through sensors is posing a big challenge in harsh cold climatic conditions. In comparison to operation under ambient weather conditions, cold region has more challenges to offer in terms of system performance and consistent reliable measurements. The drastic changes are evident in sensors measurement over a period right from the start of installation and normal routine operations till its degradation. The range of factors influencing the system performance is quite diverse in nature and always not easy to ascertain, but with the introduction of standardized design, installation, integration, and maintenance procedures the mean time between failures can be increased considerably. In the said context, cold and harsh environment plays an important role in provoking the undesirable events. This paper discusses the system performance of a custom based meteorological ice monitoring station installed in Northern Norway. The overall system design in perspective of data acquisition and logging is analyzed along with pinpointing of the degradation of the sensors and associated data validation over a period of time. The impact of the variations in climatic conditions is presented including the probable causes affecting the system performance in brief.**

Keywords*-*atmospheric icing; sensors; operation; performance; cold region.

I. INTRODUCTION

Heavy icing loads and icing events are crucial parameters for structural design in cold climatic regions. The unpredictability elements in such situations could easily cause extensive ice loading on equipment/sensors and structures within no time. Simultaneously, human activities are increasingly extending to the cold climate regions due to the availability of extensive unexplored natural resources, where, atmospheric icing not only creates human inconveniences, but also affects human activities especially in the construction industry (communication towers and ski lifts), energy distribution (power network cables and towers), maritime activities, aviation conditions on the ground, meteorological observations and wind energy power production [1]. Various structures such as, power network cables, telecommunication masts, and others have been damaged or destroyed on numerous occasions due to the added mass of ice or an increase in aerodynamic interaction leading to unacceptable movements [2]. To avoid certain situations and for a better preparedness and more reaction time for preventive measures, it is essential to predict such events involving extensive databases over longer periods of time.

Atmospheric icing on structures occurs in conditions where cooling of an air mass causes super cooling of water droplets resident in the air mass. Water droplets in the earth atmosphere can remain in the liquid state at air temperature as low as -40 ^{o} C, before spontaneous freezing occurs [3],[4]. The rate of atmospheric ice accretion on a structure is governed by two processes; the impingement of super cooled water droplets on the structure surface and surface thermodynamics, which determine what portion of the water droplets, will freeze or, on the other hand, cause melting of previously accreted ice. The systems installed under cold climate conditions to monitor atmospheric icing are subjected, directly or indirectly, to harsh environmental conditions affecting their performance and resulting in intermittent or unreliable behavior. The unpredictable nature of the climatic changes leads to uncertain behavior of the equipment, which can ultimately lead to missing project timelines, loss of significant data and excessive utilization of resources and capital. Therefore, detailed knowledge of frequency and duration of icing events, as well as, maximum ice loads are crucial parameters for the design of structures in cold regions.

To achieve this goal, an atmospheric ice monitoring station based on metrological equipment was designed within the ColdTech-RT3 project at Narvik University College, Norway. The system was installed at the mountain height, 'Fagernsfjellet (1007 m.a.s.l., 68° 25'20'' N, 17° 27'26'' E) located east from the Ofotfjorden and towards the north east from Beisfjorden on the western coast of Norway, as shown in Fig.1. The region is affected by the Gulf streams flowing across the North Atlantic Ocean, which retains the coastal area at sea level ice free. The mountain faces the open sea from the south across the SW towards the west. Air masses related to these streams are usually humid and have air temperatures favorable for atmospheric icing during the winter season (-25<T ºC <0). The objective of this paper is to analyze the performance of metrological sensors under harsh cold environment; practical issues are highlighted based on their impact on the overall system performance.

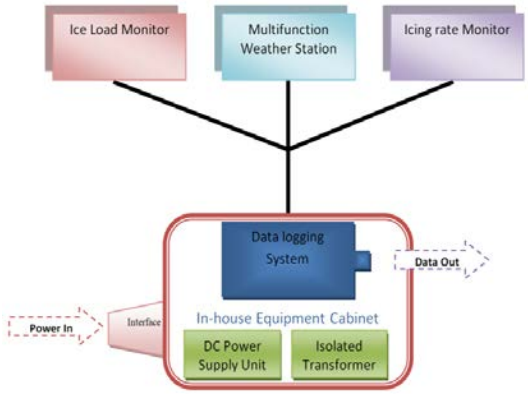
 ISBN: 978-1-61208-296-7

Figure 1. Location of HiN atmospheric icing station, Fagernsfjellet, Narvik, Norway.

This paper is arranged in three main sections. In the section two, the design of a remote icing station is briefly explained. This section is followed by the performance evaluation of this icing station (Section III). In the last section, the reasons for the station failure are evaluated based upon the experience and some technical flaws, which are planned to be avoided in the next installation.

II. DESIGN OF THE ATMOSPHERIC ICING STATION

The multi-sensor ice monitoring station used for this study was comprised of a multifunctional weather sensor (*Lambrecht – ELOS IND)* along with an ice load monitor (*Ice Monitor- Combitech*) and a (*HoloOptics T44*) icing rate monitor. A robust industrial scale data logging system Campbell Scientific (*CR1000*) was used for data-logging over a longer period of time in harsh cold conditions. The multifunctional weather station measures five parameters namely: wind speed and direction, temperature, humidity, atmospheric pressure and dew point. The ice monitor yields the accreted ice mass, whereas the HoloOptics icing rate monitor gives the icing rate. The data acquisition (*DAQ*) system and dc-power distribution/supply units (used for icing station) were housed inside the small winterized cabinet mounted inside the available facility at Fagernsfjellet, utilized for skiing and tourism.



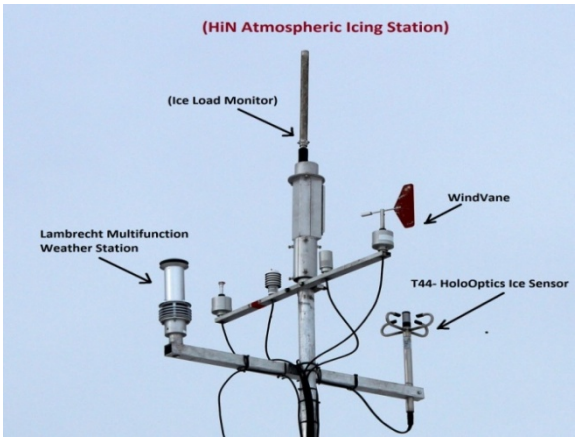


Figure 2. HiN Atmospheric icing station at Fagernsfjellet, Norway

The main power source for this station was acquired from the existing facility at the site. The data retrieval option was exercisable through RS-232 port through which data could be accessed, stored, deleted, overwritten, and appended by the vendor software interface. The overall system breakdown in terms of sensors and major elements is shown in Fig. 2.

III. ATMOSPHERIC ICING STATION PERFORMANCE

The sensors used for HiN atmospheric icing station were designed for cold climate and the performance parameter were expected to satisfy the desired outcome, but still these were affected by the cold environment with the uncertainty element inside and the sensors output was evident after the data-retrieval. Icing station was installed at Fagernesfjellet during the 2nd week of October, 2012. To have a quality check on the icing station data, icing station was planned to

 ISBN: 978-1-61208-296-7

be inspected periodically after its installation, but this task could not be accomplished several times mainly due to harsh weather conditions and associated facility's maintenance problems and logistics, which were also linked to the alternative mode of transportation to the site (*cable car and snow bikes*). The installed equipment could only be reached and tested first time four months after its installation. Initial analysis of retrieved data from the icing station showed that from the installation date the system component's operations were normal as expected, but within a few days the HoloOptics sensor malfunctioned and started to provide erroneous icing rates. Similarly, after seamless operations of one week the atmospheric temperature and relative humidity parameters from the *Lambrecht – ELOS IND* were defective. The remaining parameters of *Lambrecht – ELOS IND*, such as: wind speed, direction and dew point were satisfactory for the following two months but turned erroneous for the remaining period of operation leaving the ice load monitor as the only valid amongst the data-set recorded in the logger system. Occurrences along with related measured observations to have a vivid understanding of the system's performance in the duration span of approximately six months after installation are shown in Table 1.

TABLE 1
ATMOSPHERIC ICING STATION OPERATING PARAMETERS

S.No.	Parameter Description	Sensor	Correct Funtionality Period	Unserviceability Reported After	Last Valid Recording
1.	Ice rate	Holoptics T44	1 days	04- Months	0 (No Ice)
2.	Humdity	*Lambrecht – ELOS IND*	1-week		93.5%
3.	Atm. Temp (C)				0
4.	Dew Point Temp (C)				0
5.	Wind speed (Avg)	*Lambrecht – ELOS IND*	2 Months		6 m/sec
6.	Wind direction				NE
7.	Avg Pressure		5 Months		930
8.	Ice Load	*Ice Monitor-Combitech*	Complete Period	Nil	N/A

The equipment maintenance checks were performed in more details after data retrieval. Initial investigation at the station site disclosed that degradation of atmospheric icing station component performance was mainly the result of weather changes specifically related to the region where the combination of several climate factors negatively impacted the sensors, both in terms of the mechanical and electrical failures. Heavy ice deposition on the exposed components of sensors, power extension (interface) and frequent power break downs of the associated facility (through which the main power was supplied) caused the malfunctioning of the sensors.

Despite the fact that sensors used in the reported ice monitoring station were designed to withstand harsh cold weather conditions, the weather sensor malfunctioned unexpectedly at cold conditions. The minimum atmospheric temperature reached -25 ^{o}C at the installation site and varied considerably; it was observed that the equipment got unserviceable at 0 ^{o}C. It is worth mentioning here that the weather sensors were functional at -25 ^{o}C during this study, but analysis of atmospheric data for the last 12 hours prior to failure of the weather sensor indicated that the initial crash occurred at 0 ^{o}C, average wind speed was 6 m/sec, relative humidity was 93.5%, and most importantly, the average icing loads on ice monitor were quite high. Analysis confirms the combination of extreme weather conditions leading to heavy ice loads at the time of malfunctioning of the weather sensor. Wind direction and wind speed data for the period of 12 hours preceding the crash of the weather sensors are presented in Fig. 3.

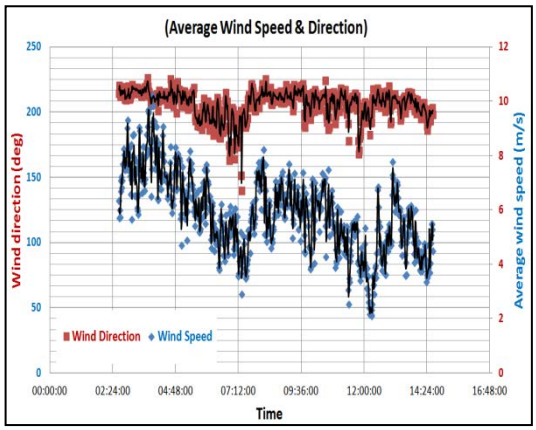


Figure 3. Wind speed and wind direction data for the seamless period of 12 hours prior to malfunctioning of the weather sensor.

It is observed from Fig.3 that wind was southwesterly at an average speed of 4 to 8 m/s. The southwesterly side of mount Fagernesfjellet faces the Beisfjord and the Ofotfjord of the Atlantic, which means that the wind is loaded with water droplets.

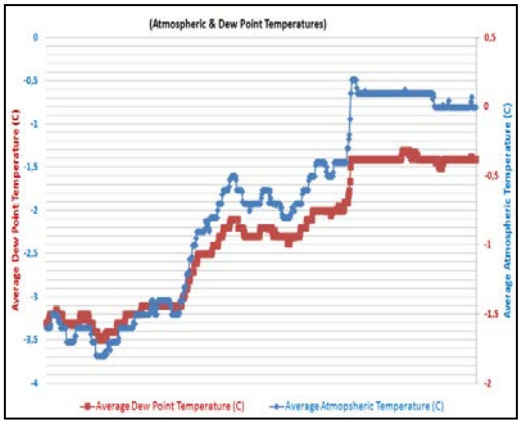


Figure 4. Atmospheric temperature and due point data for the seamless period of 12 hours prior to malfunctioning of the weather sensor.

 ISBN: 978-1-61208-296-7

Atmospheric temperature and due point data for the seamless period of 12 hours prior to malfunctioning of the weather sensor is represented in Fig. 4. It is observed that temperature was generally below freezing. It is also observed that the dew point was less than average temperature for the last few hours, which can be interpreted as a cause of freezing of the water droplets in the air mass blowing across the weather station. Those reasons coupled with the wind data indicating the presence of water vapors in the air blowing from the sea and dropping temperatures and dew point to below freezing will cause ice accretion once the droplets get in contact with freezing particles or surfaces such as the icing station or other objects in the region. Those expectations are confirmed by the available measurements of humidity and ice loads represented in Fig. 5 below.

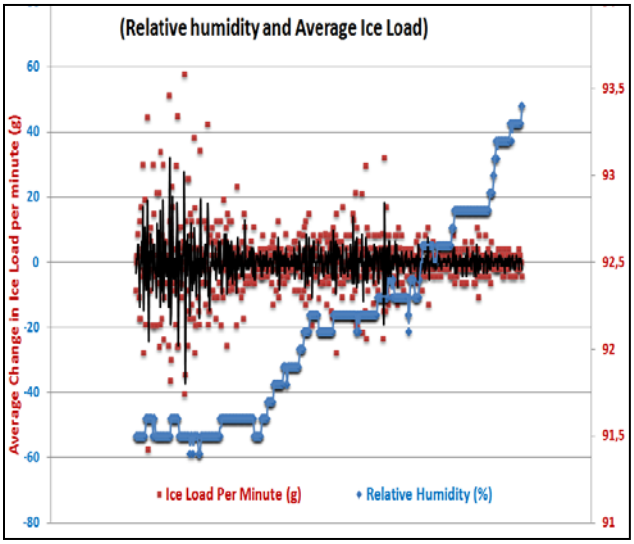


Figure 5. Relative humidity and icing load data for the seamless period of 12 hours prior to malfunctioning of the weather sensor

Fluctuations are observed in ice load measurements, but those fluctuations are in correlation with variations in humidity. The ice load readings are highest at lower humidity due to the fact that some of the humidity in the air is transformed into accreted ice on the system components.

IV. REASONS OF SYSTEM FAILURE

The dimension of operational problems faced in cold climate is quite different from the operations in normal atmospheric conditions. More often, the factors not significant at all in the normal conditions become extremely critical in cold climate regions. Investigations were carried out to track down possible reasons of the HiN icing station's components failure from operational point of view. Analysis showed that in addition to the harsh weather conditions, a combination of various design and operational aspects could also lead to the system's failure in harsh conditions. In the following, we present some noteworthy causes in this regard.

a) Intermittent Power Source

The system installed at the location takes power from the available commercial facility, where high load machines are being operated. Due to the demographic location of the site in terms of accessibility and complicated power infrastructure available in terms of maintenance, several power breakdowns have been frequently reported. The instantaneous power surge could be one reason that has affected the sensors operations.

b) Electrostatic Discharge

The electrostatic discharge phenomena could not be fully neglected in weather station breakdown. For snowstorms, temperature gradients in the ice particles produce charge separation because the concentration of H+ and OH- ions in ice increases rapidly with increasing temperature. H+ ions are much more mobile within the ice crystal than OH- ions. As a result, the colder part of an ice particle becomes positively charged, leaving the warmer part charged negatively [5]. The resulting electrostatic phenomena due to blizzard can be hazardous for the control circuitry inside the sensor module, provided the said consideration is not catered for in the design. Over and above this fact, the proper maintenance of earthing at the site becomes all the more critical in this perspective.

c) Data Links / Interfaces Winterization

Interface links between the components are data and power based. Data links might include the Ethernet/serial links with supporting routing cables or interface panels, whereas, power links have distribution panels, supplying power requirements to the computing and sensing equipment. Interface links along with power support systems have direct and/or indirect exposure to cold climatic conditions and they are under sudden transitional states, hence are most vulnerable to degradation and failure.

d) Power Cable Insulation

Electrical insulation of external power cables can be another possible cause of system failure. Many of the insulations normally used on electrical wires and cables are not compatible with colder temperatures. Cracking of the insulation exposes the conductor to the environment creating a serious hazard. This is particularly a problem for the extension cords used outdoors. Several polyvinylchloride (PVC) insulations that are commonly

 ISBN: 978-1-61208-296-7

used as electrical insulation do not withstand flexing at low temperatures, in the range of or below -30ºC, PVC insulations crack and peel off leaving exposed conductors, which can cause short circuiting or develop grounding problems making data unreliable[6],[7].

e) Material & Winterization

The sensitivity of problems encountered in cold regions is largely a function of materials used in the sensor construction and degree of stress, under which they are operated. Some materials get stronger at cold temperatures while other materials can be altered to become more cold tolerant [6],[8]. Similarly, sensor winterization can be another possible reason for this failure. Sensors must be properly winterized to make them possible to use during winter and reduce cold related wear and breakage [9].

V. CONCLUSION

Based on the experience from the operation of a multi-sensor ice monitoring station, it is evident that factors leading to sensor failure in cold climate regions are generally overlooked. Failure of the sensor components during this expedition has raised questions regarding the standards specified for the design of sensors for cold regions. Anticipated and preventive design, monitoring, and maintenance culture should be invoked as a regular practice, which could encounter the unpredictable impact of cold harsh environment upon sensor operation and performance. The problem areas identified and discussed can be utilized as guidelines for better design, integration, installation, and maintenance of sensors installed in remote cold regions.

ACKNOWLEDGEMENTS

The work reported in this paper was partially funded by the Research Council of Norway, project no. 195153/160 and partially by the consortium of the project ColdTech-Sustainable Cold Climate Technology.

REFERENCES

[1. M.S.Virk, *Numerical study of atmospheric ice accretion on various geometric cross sections.* Wind engineering, 2011. **35**(5).
2. T.G.Myers and J.P.F. Charpin., *A mathematical model for atmospheric ice accretion and water flow on a cold surface.* Heat & Mass Transfer, 2004. **47**: p. 5483-5500.
3. Kathleen.F.Jones and K.Z.Egelhofer, *Computer model of atmospheric ice accretion on transmission lines*.
4. I.J.Battan, *Cloud physics and cloud shedding*1962, New York: Doubleday and Co.
5. J.L.Shorter, *The electrification of snowstorms and sandstorms 1963*, 1963: Dept of physics Manchester College of Science & Technology.
6. Dean.R.Freitag and T. McFadden., *Introduction to Cold Regions Engineering*1996, New York: ASCE Press. 725.
7. D.V.Rosato and R.T. Schwartz, *Environmental effects on polymeric materials*. vol. 1. 1968, New York: John Wiley & Sons.
8. P.K.Dutta, *Behavior of materials at cold regions temperatures*, 1988, U S Army Corps of Engineers, COld Regions and Engineering Lab (CRREL): Hanover. p. 72.
9. D.Diemand, *Winterization and winter operation of automotive and conctruction equipment* in *Cold regions technical digest ; no. 92-1*1992, U S Army Crops of engineers, cold region reseeach and engineering lab: Hanover.

 ISBN: 978-1-61208-296-7

Early-Warning System for Machine Failures: Self-Sufficient Radio Sensor Systems for Wireless Condition Monitoring

Michael Niedermayer, Eduard Armbruster
Fraunhofer Institute for Reliability and Microintegration, Berlin, Germany
niedermayer@izm.fraunhofer.de

Rainer Wirth, Axel Haubold
GfM Gesellschaft für Maschinendiagnose mbH Berlin, Germany
wirth@maschinendiagnose.de

Stephan Benecke, Klaus Dieter Lang
BeCap
Technical University of Berlin Berlin, Germany
tib421@tu-berlin.de

***Abstract* — European machinery products are in great demand due to their reliability. This competitive advantage can be further extended by using wireless sensor network for condition monitoring. Radio sensor nodes constantly monitor the aging of machines and detect faults before the machine fails. This means that time-consuming and expensive downtime can be avoided. A sensor network for the condition monitoring of industrial machines has been developed in the publicly funded project ECoMoS. The implemented sensor nodes are able to predict machine failures until 3 months in advance.**

Keywords-Energy Harvesting; Condition Monitoring; Machine Diagnosis; Wireless Sensor Networks

I. INTRODUCTION

Systems for condition monitoring should ideally identify malfunctions before the machine fails. Until now, this kind of sensor has only been used with high-end drives for cost reasons. The situation is changing with the availability of inexpensive, self-powered wireless solutions because the burden of such condition monitoring systems drops signify-cantly. This is especially true for upgrading existing production equipment whose unplanned downtime causes considerable costs. Furthermore, wireless sensor networks open up new opportunities by distributed data acquisition and communication. This offers the chance to revolutionize measurement technology in the coming years. In the joint project 'Energy-autarkic Condition Monitoring System' (ECoMoS), the application field of self-powered wireless sensor technology has been expanded to the wireless condition monitoring of industrial plants. A lot of research has been done to prepare this development. The preceding research activities focused on the implementation of sensor nodes with small size [1,2] at minimal costs [3,4] for industrial enviroments [5,6].

The rest of the paper is organized as follows: first, we discuss relevant aspects of vibration diagnosis in Section 2. Afterwards, prototypes for energy harvesting are presented in Section 3. Section 4 describes several implementation details of the wireless sensor network. Finally, Section 6 concludes our work and gives a perspective on our future work.

II. DETECTING MACHINE DAMAGES IN ADVANCE

Several measurement categories can be checked for the condition monitoring of machines. Temperature and vibration signals are most frequently analyzed to identify malfunctions of machines. Here, we focus on condition monitoring based on acceleration sensors which measure vibrations. The vibration diagnosis has been established for early prediction of machine breakdowns. The corresponding signal analysis methods are usually based on spectral analyses of the occurring mechanical vibrations. Two types of machine diagnosis can be distinguished regarding the accuracy. One variant, the basic diagnosis utilizes methods of simple characteristics. The second variant of condition monitoring applies an in-depth diagnosis to predict machine failures several weeks in advance. The higher accuracy can be achieved with the help of diagnostic algorithms, which use knowledge of kinematic drive models in addition to the analysis of vibration measurements. Condition monitoring systems for in-depth diagnosis are based on wired acceleration sensors, which are connected with diagnosis core units. There are only several wireless system solutions to run the less precise base diagnosis of machines.

The concept in the joint project ECoMoS (Fig. 1) is based on wireless condition monitoring using wireless sensor nodes [7]. The individual sensor nodes have the intelligence to carry out an in-depth diagnosis directly at the measuring position. This makes it possible to predict, for example, bearing damage by up to three months in advance. The sensor systems can be configured via radio transceiver and regularly send the machine state to a remote base station.

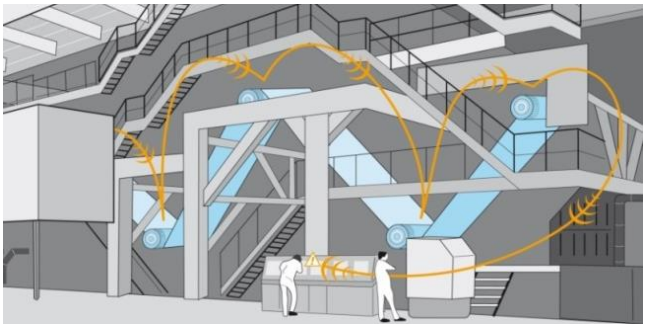


Figure 1. Condition monitoring of plants by wireless sensor networks

The complexity of the in-depth diagnosis drives the system requirements for the wireless sensor nodes to a large extent [8]. The vibration signal needs to be sampled at a frequency of 10 kilohertz and then digitized. At the same

 ISBN: 978-1-61208-296-7

time, it is required to check whether the engine speed has remained within a tolerable range of about 0.5 percent during the measurement interval. If the change in rotational speed is greater, the measurement will be discarded. Since the measurement of the engine speed is performed by a different sensor, the speed values are transmitted before and after the vibration measurement of the wireless sensor node.

The spectrum of measured time signal has to be calculated by a fast Fourier transform (FFT, [9]) to later derive the envelope spectrum of the vibration signal. The abscissa of the spectrum can be replaced from sampling frequency to the fraction of the basic rotational speed to get the order spectrum. The critical multiples of the basic rotational speed are identified by a significance analysis. In a further step, it is checked whether these coincide with typical damage patterns. If errors are detected, e.g., irregularity on the inner ring of the roller bearing SKF 6309, the results of the diagnosis are transmitted in the form of feature vectors, comprising error code, kinematic frequency, significance level, and magnitude.

For the implementation of wireless sensor nodes that can perform this type of machine diagnostics directly at the measuring position, different circuit designs were evaluated regarding their energy efficiency. Some Digital Signal Processors (DSP) were capable of performing the algorithms for in-depth diagnosis more energy-efficiently than 16-bit microcontrollers, e.g., MSP430 from Texas Instruments, or 32-bit microcontrollers, e.g., AVR32 from Atmel. Hence the power consumption of the processor types could be compared for a number of FFT cycles with 1024 samples at a clock frequency of 1 MHz (Fig. 2). While the DSP based system architectures require significantly higher peak currents, the execution time is considerably shorter. This results from the special hardware support for filters and FFT calculations.

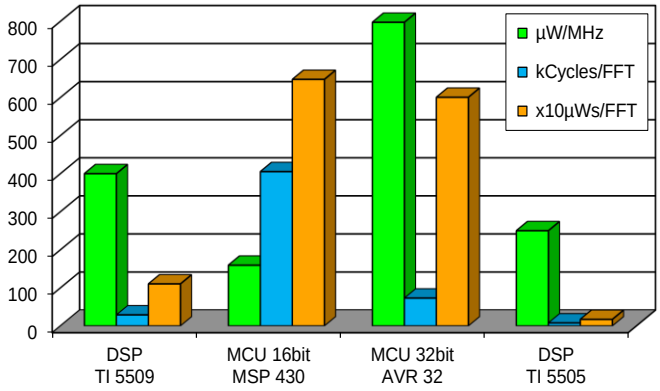


Figure 2. Comparison of different system architectures [10]

The analysis of analog circuit elements for data acquisition showed that the use of a DSP system architecture is the best solution to minimize the power consumption because the data from the acceleration sensor can be digitized immediately while the necessary filtering and scaling is left to the DSP. Another advantage of this approach is the increased flexibility in the evaluation of the measurement. It is thus possible to generate customized reports for specific measurement positions by parameterization with moderate effort. Fig. 3 summarizes the system architecture of the ECoMoS sensor nodes.

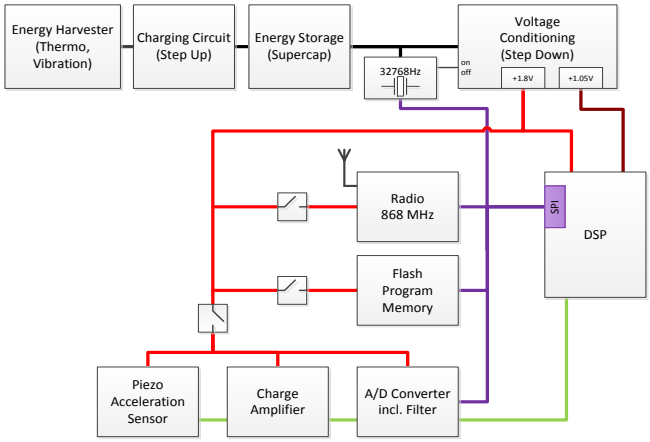


Figure 3. System architecture of the radio sensor nodes [10]

The selection of the appropriate sensor depended on accuracy of the measurement range, size, supply voltage, and power consumption. It was found that a piezoelectric acceleration sensor from IMI Sensors with charge output provides the preferred properties for the target application. Thus, a charge amplifier has been added to the input stage on the sensor node. This makes it possible to sample the machine vibrations with minimal energy consumption. Furthermore, the duration of the transient can be considerably shortened by a special discharge stage. This would not be possible for sensors with integrated amplifier.

III. ELECTRIC ENERGY FROM VIBRATIONS AND TEMPERATURE DIFFERENCES

In addition to precise sensor, efficient data processing, and wireless connectivity, the power supply belongs to the key components for the development of maintenance-free wireless sensor nodes. The operating time of the wireless sensor nodes with batteries – depending on the duty cycle of the in-depth diagnoses – can reach several years by application of low-power components in combination with an optimized power management. However, there are several concerns regarding battery-powered wireless sensors especially in industrial environments. The users prefer solutions energy harvesting whenever possible. While solar cells can only be used in very clean industrial surroundings with little dust, thermal generators and vibration transducer provide a good way to convert the energy from the environment. This eliminates any effort for battery replacement and increases the operational reliability.

There are many different types of vibrations, which can be used for powering wireless sensor nodes in industrial environments. However, rigid structures should be chosen as measurement position for in-depth diagnosis of machines. This improves the precision of the diagnosis results but the magnitudes of vibrations on these locations are typically below 2m/s² (Fig. 4).

 ISBN: 978-1-61208-296-7

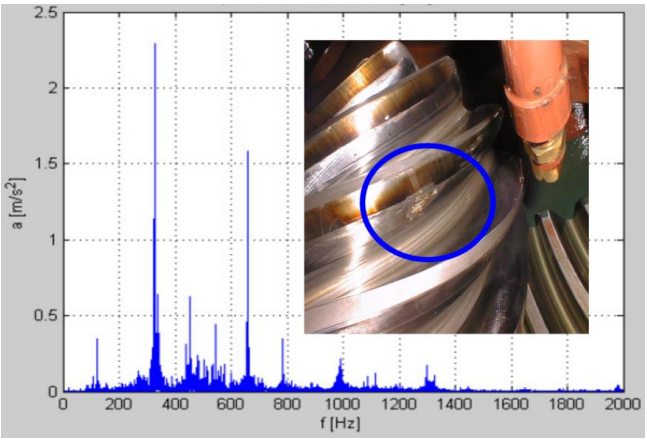


Figure 4. Exemplary acceleration spectrum of a typical drive

Most commercial vibration transducers are designed for a single resonance frequency between 50 and 120 Hertz. This means that such a narrowband converter cannot be used on each machine, since the efficiency massively decreases whenever the vibrations are not in the vicinity of the resonance frequency.

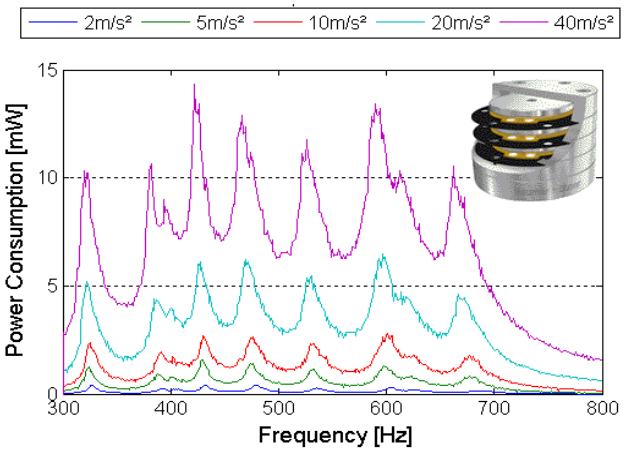


Figure 5. Design of a broadbane vibration transducer by stacking piezo elements [10]

A broadband vibration transducer (Fig. 5) has been developed in the ECoMoS project by skillful stacking piezoelements. The first prototypes of the vibration energy converter (Fig. 6) were tested with a tunable Shaker. It was found that the vibrations in the range of 300 to 800 Hertz provided enough energy to supply the radio sensor electronics. Such broadband transducers can be used more universally.

Significant temperature differences can often be found in industrial plants. Different drives have been examined under operational conditions regarding their temperature distribution (Fig. 7a) to develop very compact wireless sensor nodes with integrated thermal converter. It was considered what temperature differences occur directly on the thermocouple for typical measuring points by using a special test setup consisting of a reference heat sink and thermal converters with embedded temperature sensor. A temperature difference of only 3 ° C at the thermocouple was required to deliver enough energy for the power consumption of wireless sensor nodes for condition monitoring.

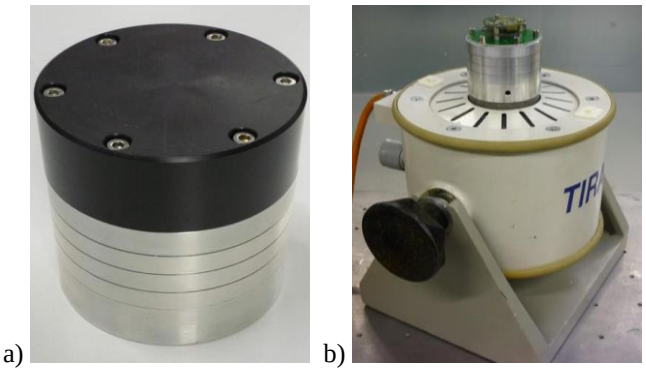


Figure 6. a) Implemented prototype of the vibration harvester, b) Vibration transducer characterisation, source: Baumer

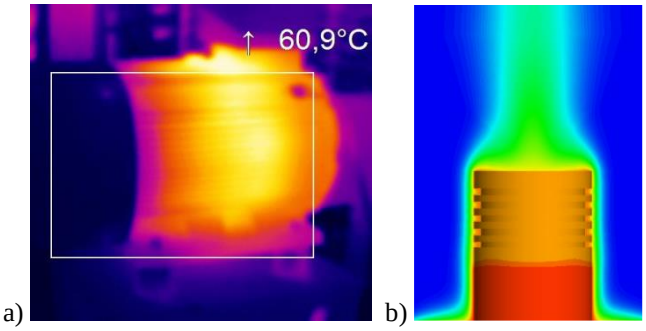


Figure 7. a) Temperature distribution during machine operations, b) Simulation of temperature distribution

The knowledge of the temperature distribution at the measuring points was applied in simulations to evaluate different assembly concepts (Fig. 7b). A robust housing with a diameter of only 4 centimeters was designed so that the hot and cold sides of the thermal converter are coupled particularly well with the industrial environment, without reducing the accuracy of the high-precision acceleration sensor (Fig. 8).

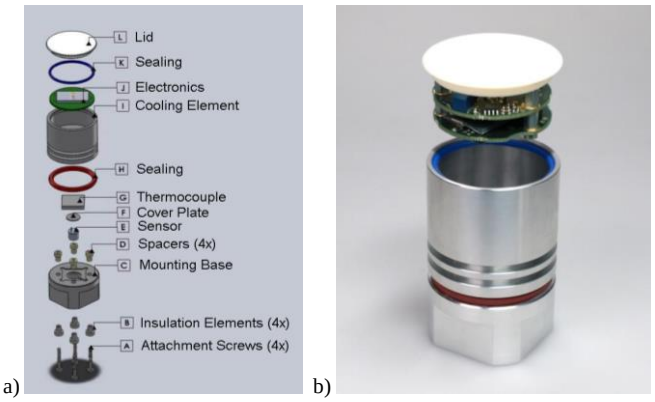


Figure 8. a) Assembly concept of the radio sensor system with integrated thermal energy havesting, b) Implemented prototype

 ISBN: 978-1-61208-296-7

A design tool has been implemented to facilitate the future development of wireless sensor systems with embedded energy harvesting. This software provides the system specific parameters from component library. The parameters can be adjusted by the hardware designer to layout the power supply of self-powered wireless sensor nodes. After selection of the conversion method, the type of physical size of the harvesting unit can be suitably configured. In addition, parasitic coupling effects are taken into account. In the example of the thermal converter, this applies to additional layers, (e.g., thermal paste and mounting adapter), which have to be taken into account for their impact on the energy efficiency (Fig. 9).

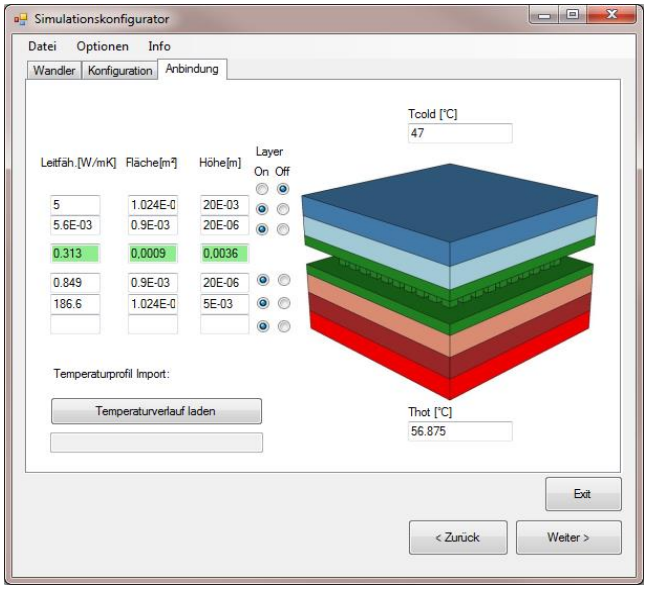


Figure 9. Selection of the energy converter and configuration

Following the energy conversion chain, the parameters for the voltage conditioning and energy storage can be queried. It is possible to select predefined voltage transformers. During simulation the input or output impedance at the operation point as well as the converted output voltage or current level are considered together with non-ideal effects of energy buffers such as leakage current and a parasitic internal resistance. In the input mask for configuration of the load, the clock frequency of the processor is to be set in order to determine the computation time for the individual algorithms. This approach allows for a first estimation of the energy performance of the wireless sensor node to determine the size of the power conversion stage and energy buffer so that particularly compact sizes can be realized inexpensively. The use of System-C for system modeling is also planned to investigate more complex algorithms in terms of energy requirements.

IV. INSTALLATION AND MAINTANENCE PER WIRELESS NETWORK

In production halls, there are many sources of interference that can affect radio transmission considerably. In the early days of radio communication, many problems regarding robustness of radio links occur in the industrial sector. This led initially to a high skepticism about wireless sensors, because the robustness of data transmission in industrial environments is particularly important. Meanwhile, there are various solutions to ensure robust radio communications. In particular, procedures at the protocol level of digital radio communication can reconstruct the correct data even when data erasure of multiple bits occurs. Therefore, the problem of robust data transmission is well under control in most cases, although the increase in energy efficiency of robust radio communication still remains an important research topic. An energy-efficient TDMA-based MAC protocol [11] has been applied. In this single-hop network, all sensor nodes receive a time schedule about the next communication interval from the base station.

The installation of wireless sensors can now be done quite comfortably. After attaching to a machine, sensor nodes identify existing wireless networks and log on. The operating parameters of the sensor systems, such as duty cycles, can be adjusted via the radio base station. A software update via radio link is also possible if new damage patterns or changes in communication protocols require reprograming the wireless sensor nodes. Thus, the user does not need to trigger a time-consuming firmware update for each individual sensor nodes via cable. Such a wired reprograming of sensor node can be very expensive especially for measurement positions, which are hard to reach.

For the preparation of the field tests, the prototypes were first tested in the laboratory by a tunable Shaker, which simulated the vibration profile of broken machines. The direct field trials were carried out in the harsh environment of a paper mill under realistic operating conditions in order to obtain more meaningful test results.

Figure 10. Paper plant in Spremberg [10]

On the site of the former gas conglomerate, Schwarze Pumpe, between Cottbus and Dresden, where earlier lignite was refined, paper production has been in full swing since April 2005. 170 million euros were invested to build the Spremberg mill. With a capacity of 330,000 tons of base paper per year, corrugated board is produced from 100 percent recycled paper. The preparation of a 1000 meter long and 5.4 meter wide paper web requires just a minute. Only one hour of downtime would cost 5,000 euros. Wireless technology significantly reduces material and installation costs for copper wire and cable channels about 100 meters

 ISBN: 978-1-61208-296-7

total length of the paper machine (Fig. 10), that the initial investment for a wireless condition monitoring system offers a very short payback time.

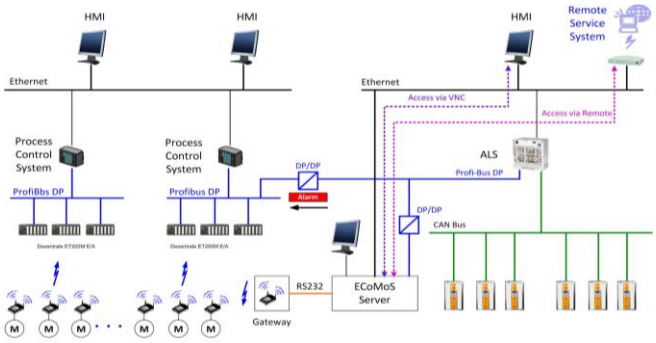


Figure 11. Integration of the ECoMoS sensor network into the process control system of the paper plant [10]

The fusion of data from the wireless sensor nodes is provided by the base station, the so-called ECoMoS server, which is connected by field bus with the control center of the paper machine (Fig. 11). During the installation of sensor node as well as the exchange of production components the kinematic parameters for the in-depth diagnosis has to be updated at the ECoMoS server. The ECoMoS server also receives the rotation speeds of the individual drives in the paper machine and transmits this data to the corresponding sensor nodes ted to the server ECoMoS. This procedure is done before and after each measuring interval.

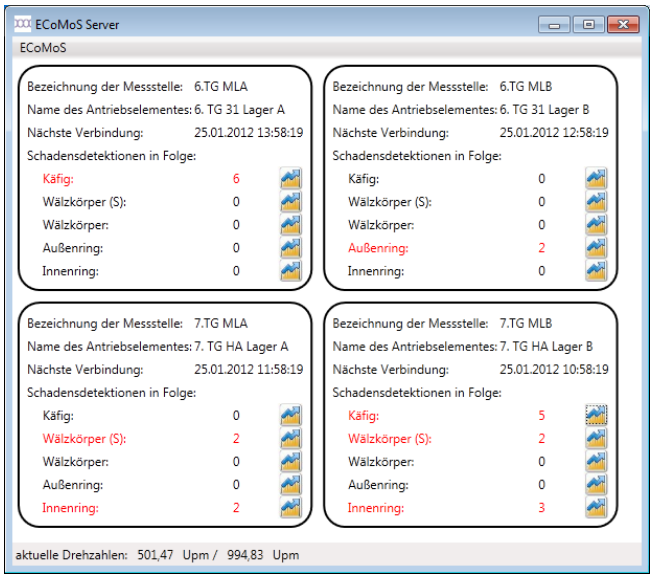


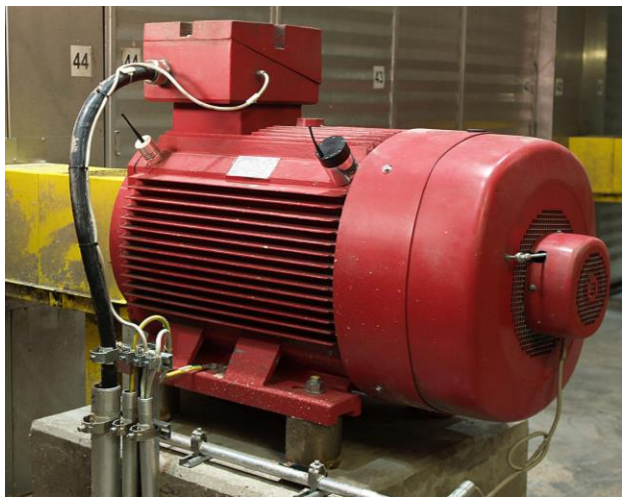
Figure 12. Result visualization of the in-depth diagnosis

If a damage pattern is detected, each sensor node can send the raw data via the base station and the ECoMoS server to the control center of the paper plant (Fig. 12). So, experts can examine the characteristics of serious errors in critical cases before costly actions for damage repair are initiated. Often, it is sufficient to replace a damaged bearing during the regular maintenance interval.

V. SUCCESSFUL PRACTICAL TEST IN THE PAPER FACTORY

In the first prototype generation, the system parts sensor, data processing, radio interface and power supply had been designed as modules to be tested in the paper mill separately. The second prototype generation as a compact complete system was implemented in 2012. Currently, the last work on the wireless protocol and the software for the ECoMoS server ECoMoS is nearly finished. Firstly, four measuring positions in the paper mills were equipped with the ECoMoS wireless sensor nodes to demonstrate a successful project completion (Fig. 13). Now, this sensor network will be progressively extended to the full configuration with 664 sensor nodes.

(a)



(b)

Figure 13. a) Field tests of the second prototype generation, b) ECoMoS-Server in operation

The wireless sensor network, developed ECoMoS project, can be used to detect wear of machines and plants that have drives, which are at least temporarily operated at a constant speed. There are many such manufacturing facilities. If later low-cost wireless sensor systems are available, condition monitoring will be used for additional

 ISBN: 978-1-61208-296-7

rotating parts such as fans, which are not currently monitored due to cost reasons. The bill of material for such sensor node ranges from 10 to 300 euros depending on the requirements and environment. The cost of an entire sensor network, however, is mainly influenced by the production quantity. The development costs dominate the overall cost in niche applications. For a wide range of applications, however, such wireless sensor nodes will be produced in large quantities, so that the design and software costs only account for a small proportion of total costs.

VI. CONCLUSION

The aim of the joint project ECoMoS was to develop a wireless sensor network for the condition monitoring of industrial machines. The wide range of expertise in the fields of measurement technology, wireless communication, energy harvesting, module integration and machinery diagnosis were necessary to demonstrate the opportunities of modern wireless sensors for condition monitoring. The installation of sensor networks within plants provides the basis for advanced concepts in condition monitoring. Wireless solutions for machine diagnosis make it possible to reduce installation costs while collecting and evaluating more data on the individual sensors.

As such systems become available, the leverage effect for the construction of machines and plants becomes hard to predict. Such sensor networks will be connected with cloud services, so that the machine diagnosis can be supported by external service providers or the system manufacturer. The manufacturer would then be in a much better position to assess the reliability of their products. It can be observed, for instance, whether the construction of machine components needs to be improved, or whether it is possible to reduce costs, without impairing the reliability.

ACKNOWLEDGMENT

The joint project ECoMoS was funded by the Federal Ministry of Education and Research (BMBF) and managedwith support of VDI / VDE Innovation + Technik GmbH in Berlin. The project results has achieved by the cooperation of the companies IMC Meßgeräte, Elbau, GfM Gesellschaft für Maschinen-diagnose, GE Energy, Baumer, EnOcean, Hamburger Rieger together with the Fraunhofer IZM and the Technical University Berlin.

REFERENCES

[1] M. Niedermayer, R. Thomasius, D. Polityko, K. Schrank, S. Guttowski, W. John, and H. Reichl, “Miniaturization Platform for Wireless Sensor Nodes Based on 3D-Packaging Technologies,” Proc. of IEEE Int. Conf. IPSN/SPOTS, Nashville, TN, Apr. 2006, pp 391–398.

[2] M. Niedermayer, J. Hefer, and S. Guttowski, “Technology-Oriented Design of Radio Sensor Nodes,” in Smart Systems Integration and Reliability, Ed. B. Michel, Goldenbogen, ISBN: 3540238670, Dresden, Germany, 2010

[3] M. Niedermayer, C. Richter, J. Hefer, S. Guttowski, and H. Reichl, “SENESCOPE: A Design Tool for Cost Optimization of Wireless Sensor Nodes,” Proc. of IEEE Int. Conf. IPSN/SPOTS, San Francisco, CA, Apr. 2009, pp. 313–324.

[4] M. Niedermayer, “Commercialisation and Application Driven Economic Viability of Sensor Technology,” in Comprehensive Materials Processing, Ed. S. Hashmi, Elsevier Science Ltd, ISBN: 0080965326, Oxford, UK, 2013

[5] E. Hohwieler, E. Uhlmann, M. Niedermayer, and B. Hirsch: “Self-organizing Production with Distributed Intelligence,” Proc. 2nd Int. Symp. on Innovative Production Machines and Systems, Ischia, Italy, May 2009, pp. 163–167.

[6] M. Niedermayer, H. Scholtz, T. Bonim, S. Guttowski, and K. D. Lang, “Robust 3D Radio Sensor Systems with Embedded Active and Passive Components for Industrial Applications,” Proc. Int. Conf. on Sensors and Its Applications SIA, Jeju Island, Korea, Nov. 2012, pp. 336–341.

[7] E. Kravcenko, M. Niedermayer, S. Guttowski, N. Nissen, S. Benecke, A. Middendorf, and H. Reichl: “A System-oriented Approach for Modeling Energy Harvesting Devices in Wireless Sensor-Modules, ” Proc. of 4th Int. Conf. on Sensor Technologies and Applications SENSORCOMM, Venice, Italy, Jul. 2010, pp. 305–310.

[8] M. Niedermayer, “Cost-Driven Design of Smart Micro Systems,” Artech House Publishers, ISBN: 1608070840, Norwood, MA, U.S. 2012

[9] E. O. Brigham, “The Fast Fourier Transform,” Prentice-Hall, New York, 2002

[10] R. Wirth, et. al., “Self-Sufficient Condition Monitoring System ECoMoS”, Grant 16SV3371-16SV3377, Project Report, TIB Hanover, 2013

[11] S. C. Ergen, P. Varaiya, “TDMA scheduling algorithms for wireless sensor networks”, Springer Wireless Network, Vol. 16, 2010, pp. 985–997.

 ISBN: 978-1-61208-296-7

Animal Sensor Networks: Animal Welfare Under Arctic Conditions

Mohamad Y. Mustafa
Department of Technology
Narvik University College
Narvik, Norway
myfm@hin.no

Inger Hansen, Svein Eilertsen
Bioforsk Nord Tjøtta
Tjøtta, Norway
inger.hansen@bioforsk.no
svein.eilertsen@bioforsk.no

***Abstract*—Animal tracking, which can be efficiently performed using wireless sensor networks (WSNs) is an important tool that provides useful information about animals' activities. It is also fundamental for environmental research. Under arctic conditions, WSNs face many challenges, such aslogistics costs, accessibility, danger of animals that might eat or destroy the equipment, effect of low temperatures on components, battery durability (which is drastically affected by low temperatures), effect of snow and ice on antennas and moving parts and continuous needs for calibration and adjustment due to extreme seasonal changes. However, they are a very good tool to overcome accessibility of semi-domestic animals under harsh weather conditions. One possible application of WSNs is the identification of ownership of newly born reindeer, which will be presented in this work as an example on arctic application.**

Keywords-Wireless Sensor Networks (WSN); Animal Sensor Networks (ASN); RFID; Wi-Fi Enabled RFID; ZigBee

I. INTRODUCTION

The use of WSN to monitor animal movement and behaviour is increasing rapidly. WSNs can be of great value in managing the animals and improving their productivity and impact on the environment. They are also very useful for animal owners and researchers for acquiring useful data, which can be used for the welfare of the animals.

Under arctic conditions, the environment imposes certain challenges on the network due to harsh weather conditions, such as logistics costs, accessibility, danger of animals that might eat or destroy the equipment, effect of low temperatures on components, battery durability (which is drastically affected by low temperatures), effect of snow and ice on antennas and moving parts and continuous needs for calibration and adjustment due to extreme seasonal changes. These factors are usually hard to predict by developers and will only appear through actual implementation of WSN strategies, which are much more closely integrated to the environment than other systems.

On the other hand, the harsh conditions in the arctic region necessitate the implementation of WSNs to overcome the problem of accessibility and reduce logistics and cost for acquiring necessary data about animals and the environment. This is consistent with the fact that science contributions from the computer science community have always been application-driven, which remains an essential philosophy for (WSN) community. Those challenges and contributions are presented within the Animal Sensor Networks (ASN) project.

The ASN project is part of The Botnia-Atlantica programme, which is a cross-border cooperation programme, intended to co-fund projects within the Botnia-Atlantica area. The programme is one of several European territorial cooperation programmes, co-funded by the European Regional Development Fund. The programme covers a larger area that includes the regions Ostrobothnia, Central Ostrobothnia and Satakunta in Finland, Västerbotten and Västernorrland as well as the municipality of Nordanstig in Sweden and the Nordland region in Norway, Fig. 1. The regions Southern Ostrobothnia in Finland and Gävleborg (except Nordanstig) in Sweden participate as adjacent areas in the programme [1]. Within this region, several research groups are working in the field of animal sensor networks. Some examples are Umeå University and the project "Digital Zoo" where a large scale testbed for wireless video sensor networks was built together with Lycksele zoo [2].

The Swedish university of agricultural science (SLU) and Centria in Österbotten, Finland, have already done research related to farming. The purpose of this project is to conduct applied research in the area of ASN so as to strengthen the collaboration between these groups and establish a knowledge cluster in this field. Specifically, the project will develop systems to collect information about individual animals in large areas.

Figure 1. Interest area for the EU Botnia-Atlantica project [1].

This research will be applied for the tourism sector in the test-bed that has been developed in Lycksele Zoo [2]. The technology will also be applied for farming applications both in the SLUs experimental farm in Umeå and at Centria, Finland [3].

So far, a competence cluster around WSNs for animals has been established in the region and one test-bed and reference deployment for sensor networks applied in tourism application has been established in Lycksele zoo, together with two test-beds for sensor networks applied in farming applications has been established. Furthermore, new sensor network hardware and software platforms for collecting information from animals have been developed and long-term tests have been carried out at the test-beds for wireless sensor networks. The goal is to make the system easy to deploy and maintain. This is of high importance to create business opportunities from the technology. The project also tries to verify new innovative ideas and applications based on the technology developed in the project and develop animal tracking algorithms based on multi-modal sensor fusion. The project also aims at developing new sensor network platforms for collecting information from animals. One platform of particular interest is smart ear tags containing sensors, microcontroller, wireless transceiver, batteries and solar panel for energy harvesting. This platform can be used to collect information from animals both in farming and in tourism applications.

Tracking of animal's movement is a fundamental tool to provide good and efficient management of these animals. The study of animal's movement is also a foundation for environmental research. One possible application of this approach is the identification of ownership of newly born reindeer which will be presented in this work as a study case. ASN also have potential applications in animal monitoring for farming improvement because with this system the farmers can systematically optimise their animal business [3].

As far as the research in Norway is concerned, the research teams at Bioforsk Tjøtta and Narvik University College will help to develop the cross-border network between research groups in this field within the region Botnia-Atlantica. Other Norwegian companies are also involved in this project such as Tele Track, NoFence and Biocontrol/Os ID, all of which may contribute to the further development of electronic monitoring for grazing animals [4].

In this paper, an overview of significant WSN technologies for animal welfare will be discussed. The major WSN technologies, such as RFID, GPS, Wi- Fi and Wi-Fi enabled RFID will be considered with some details; other technologies, such as ZigBee, Bluetooth and GSM/GPRS will be briefly compared to the WiFi technology; however, the description of those technologies will take a qualitative approach and will not go into technical details. An application example of WSN for animal welfare will be presented to explain the use of this technology for ownership recognition of newborn reindeer calves. This application implies the implementation of localization algorithms, which could be material for further work.

II. OVERVIEW OF SIGNIFICANT WSN TECHNOLOGIES FOR ANIMAL WELFARE

WSNs have been used extensively for the welfare of animals. Various examples of their applications are reported in the open literature. Ranjan et al. [5] created a wireless sensor network for wildlife monitoring; in their system they utilised specially designed collars with sensor nodes, which are attached to the necks of wild animals. These collars collect data about the desired parameters from the vicinity of the animal and transmit it to the base station. This information is used by wildlife scientists for habitat monitoring and behavioural habits of the monitored animals.

Kwong [6] presented a wireless communication solution that fulfills the requirement for intensive condition monitoring of individual animals, aggregation and timely reporting of data to the farm manager designed for both loose house dairy cattle and free ranging beef cattle. The design target utilises inexpensive, low power consumption sensor nodes as the base elements of a data gathering and communication infrastructure. Laboratory experiments and farm trials have been carried out to evaluate the performance of the platform communication protocol. The results of experiments demonstrate that the platform performs efficiently while conforming to the limitations associated with WSN implementations.

In the following sections, some of the potential WSNs technologies will be considered. A study case based on Wi-Fi enabled RFID will be presented, as an example.

A. RFID

Radio Frequency IDentification (RFID) is widely used for electronic identification and tracking. RFID offers substantial advantages for businesses allowing automatic inventory and tracking on the supply chain. This technology plays a key role in pervasive networks and services [7]. Data can be stored and remotely retrieved on RFID tags, enabling real-time identification of devices and users. However, the usage of RFID could be hugely optimized if identification information was linked to location [8]. RFID networks are composed of three different entities, RFID tags, readers, and servers, as represented in Fig. 2.

All RFID tags use radio frequency energy to communicate with the readers. However, the method of powering the tags varies. An active tag embeds an internal battery which continuously powers it and its RF

 ISBN: 978-1-61208-296-7

communication circuitry. Readers can thus transmit very low-level signals, and the tag can reply with high-level signals. An active tag can also have additional functionalities such as memory, and a sensor, or a cryptography module. On the other hand, a passive tag has no internal power supply.

Generally, it backscatters the carrier signal received from a reader. Passive tags have a smaller size and are cheaper than active tags, but have very limited functionalities. The last type of RFID tags is semi passive tags. These tags communicate with the readers like passive tags but they embed an internal battery that constantly powers their internal circuitry.

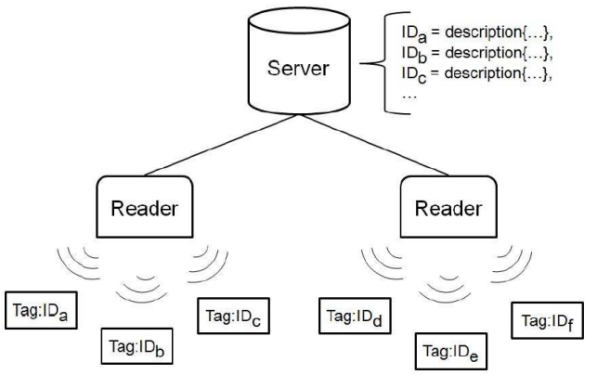


Figure 2. Structure of a typical RFID system [8].

RFID readers have two interfaces. The first one is an RF interface that communicates with the tags in their read range in order to retrieve tags' identities. The second one is a communication interface, generally IEEE 802.11 or 802.3, for communicating with the servers. Finally, one or several servers constitute the third part of an RFID system. They collect tags' identities sent by the reader and perform calculation such as applying a localization method.

RFID systems can be classified in two main categories in accordance with their usage: monitoring and authorizing. Classic applications are livestock or people monitoring. The second class includes RFID systems where the tags are not permanently attached to entities. Typical usages of authorizing RFID systems are access control in a building where tags are embedded inside cards or keys [9].

B. GPS

The Global Positioning System (GPS) has traditionally underpinned the development of ubiquitous positioning solutions. This stems from the increasing recognition across society that under ideal operational conditions, GPS can meet the majority of attributes of ubiquitous positioning system i.e. accuracy, reliability and availability. However, it is also widely acknowledged that in certain environments these performance attributes can quickly deteriorate to unacceptable levels, with often only weak GPS signals available Wi-Fi [10].

C. Wireless-Fidelity (Wi-Fi)

Wi-Fi signals are different in this regard. Over the last few years, millions of Wi-Fi access points have been have been established to facilitate more efficient and flexible access to data and communications around the world. However, these access points repeatedly broadcast a signal the strength of which can be measured and used in algorithms to locate a Wi-Fi enabled device. Unfortunately, the positioning capabilities of Wi-Fi are weakened by the fact that positioning performance is highly dependent on the strength of the signal received from nearby access points, which are affected by a diversity of uncontrollable environmental effects, e.g., people, building material, etc. Despite this significant limitation Wi-Fi signals continue to be an attractive alternative positioning technology due to the growing number of Wi-Fi enabled mobile devices being produced (over 360 million Wi-Fi handsets projected to be sold annually by 2011) combined with the rapid growth in Wi-Fi access points being installed globally (shipments of consumer oriented Wi-Fi access points are expected to grow from 6 million in 2008 to 88 million in 2013) [10].

D. Wi-Fi Enabled RFID

Wi-Fi-enabled RFID is commonly used for location-based services that track objects in a specific physical context, like children in a theme park, cars in a parking lot, equipment in a manufacturing plant, etc. It is considered a more accurate system than a traditional RFID network for determining the location of tagged objects. A regular RFID system can give what is called the "choke point" location, or zone-based location, meaning the location of the tag is known only in relation to the reader detecting its presence. A Wi-Fi network on the other hand can determine the precise *x,y* coordinates of a tag using triangulation methods, similar to how GPS works [11].

In these RFID systems, tags are Wi-Fi devices, and Wi-Fi access points function as readers. These tags can have an accuracy of 3-5 meters when determining the location and a large number of tags can communicate simultaneously with a single access point without choking the Wi-Fi network. These tags are generally capable of bi-directional data exchange [12].

The MAC (Media Access Control) Address uniquely identifies each node of a network and serves as the unique identifier on the tag, while the location software determines the specific coordinates. Additional application software can transform coordinates in terms that are meaningful to a user by matching them to specific locations. A Wi-Fi application then can identify the tag and its location. Some tags may contain additional data derived from integrated sensors [13]. This technology has been used and proven in many applications [11].

To simplify the manner of operation of the Wi-Fi-based active RFID tags, they simply send a wireless 802.11 signal at a regular interval, the signal is received by the Wi-Fi

 ISBN: 978-1-61208-296-7

access points or location receivers, and is sent to a Location Engine which determines the tag's location, and sends it to the visibility software. The visibility software uses location data to determine the exact location of the RFID tag [14]. The Wi-Fi (IEEE802.11WLAN) can support many devices (max. 128) within one network, it can transmit data up to 30m distance, needs more power, and cost is more than Bluetooth or ZigBee [15].

E. Wi-Fi vs. other wireless technologies

Table I compares Wi-Fi technology to ZigBee, Bluetooth and GSM/GPRS technologies. Obviously the ZigBee is more efficient in terms of reliability, cost, size and battery life compared to Wi-Fi, while Wi-Fi is recognised for its flexibility, speed and availability.

TABLE I. COMPARISON OF WI-FI TECHNOLOGY WITH OTHER WIRELESS TECHNOLOGIES [16].

Market Name/ Standard	ZigBee (802.15.4)/(Helicomm)	WI-Fi (802.11b)	Bluetooth (802.15.1)	GSM/GPRS
Application Focus	Monitoring and Control	Web, Email, Video	Cable Replacement	Wide Area Voice and Data
Success Metrics	Reliability, Power, Cost	Speed, Flexibility	Cost, Convenience	Reach, Quality
Battery Life (days)	100–1000+	.5–5	1–7	1–7
Network Size	Unlimited (2^{64})	32	7	1
Bandwidth (KB/s)	20–250	11,000+	720	64–128+
Transmission Range (meters)	1–100+	1–100	1–10+	1000+

Hence, for an application such as the identification of calf and mother reindeer, the existing Wi-Fi network at the premises can be used and the tags can be used for this particular purpose for a limited time each year which reduces the cost of the equipment needed, a further economic study of the required equipment should however, be performed.

III. APPLICATION EXAMPLE

Reindeer herding is the major traditional economic activity of the Saami people, who are the indigenous people of Finnmark, the northern most, largest and least populated county in Norway. Herds of Reindeer of mixed age and sex varying in size from 100 to 10,000 animals are maintained on natural mountain pasture all year round and typically are moved between coastal summer pastures and inland winter pastures. Identifying the ownership of new-born calves is a major problem for reindeer herders. Recognising and matching the reindeer calves to the different mothers is a two-step process. First the animals are caught in a small pen. All female reindeer get individual numbers sprayed on the skin at each side of the animal. Each owner uses one specific colour. Calves get a collar each around their neck, with a unique number plate hanging on it. This procedure is stressful for the animals. Thereafter, the animals are released to larger pens to calm down. Several reindeer owners and herders start observing the reindeer with binoculars in order to identify/match which reindeer calf (number plate) is following which mothers (sprayed number). Thereafter, the reindeer owners compare the different observations to guaranty that the correct mother is matched to her calf.

The next step is to gather the reindeer herd again in a small pen, catch the calves, remove the number plate and replace it by an ear cutting or a tag. This handling procedure is also stressful to the animals.

In this study, a technique based on WSNs to identify the ownership of the calves is presented. It is proposed to equip the newly born calves with an active RFID tag, facilitated with receivers and gateways based on Wi-Fi network, usually available and accessible at the identification site and monitor the movement of the herd in a confined space, such as a fenced pen, covered by a grid of signal receivers corresponding to the transmitter device attached to the reindeer, which basically tracks the movement of the tags. This process is performed in a few days per year, hence the Wi-Fi network can be utilised for other objectives for the rest of the year. A suitable algorithm has to be used to analyse the gathered data and recognise the pairs of tags which keep close together most of the time, which are the mother and calf reindeer.

Being a wild animal species, the reindeers tend to be more cautious than domestic animals; hence they tend to avoid confined spaces and close objects mounted by humans, which makes the use of passive RFIDs more difficult due to their low range, because the receivers have to be placed close to the RFIDs. This will demand the attachment of active RFIDs to both mothers and calves.

In this procedure, the pen where the animals are to be gathered for tagging is facilitated with a Wi-Fi network, a group of Wi-Fi routers mounted on posts, as shown in Fig. 3.

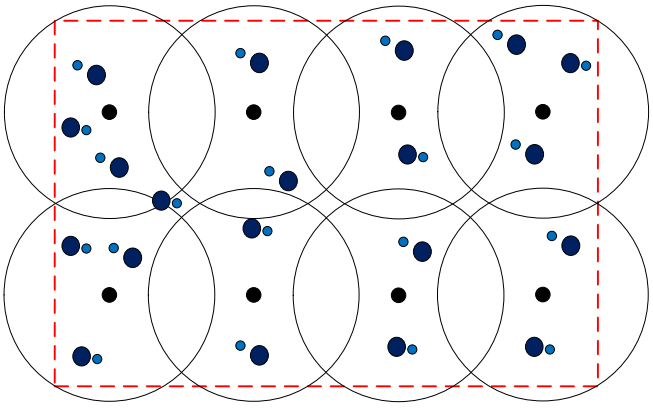


Figure 3. Animal pen of area 40 x 80 m^2 (dashed rectangle) covered by Wi-Fi network, represented by the large circles, the Wi-Fi routers are the black dots at the centres of the larger circles.

 ISBN: 978-1-61208-296-7

Large and small pairs of coloured circles represent mother and calf. The schematic represents possible configuration of the Wi-Fi coverage, which covers the whole pen, with minor pockets of uncovered areas, which have to be kept to a minimum.

The locations or the RFID tags are gathered in a computer programme and analysed. However, the proposed method suffers from a few limitations pertaining to the number of tags that can be covered using one Wi-Fi reader and coverage of the pen area, which could create further problems pertaining to transmission, repeatability and accuracy. Nevertheless, those limitations can be overcome using proper antennas and larger number of Wi-Fi readers.

In the proposed set up, tags are Wi-Fi devices, which can have an accuracy of 3-5 meters. An advantage of this system is that a large number of tags can communicate simultaneously with a single access point without choking the Wi-Fi network, while a traditional RFID technology would have required hundreds of readers in order to pinpoint the location of a single tag with the same precision as the Wi-Fi network.

IV. CONCLUSION

In this paper, the background of the animal sensor network project, which is part of the EU Botnia-Atlantica programme, has been presented. Various significant (WSN) technologies have been presented and an example application for animal welfare has been described.

WSNs face many challenges under arctic conditions, but at the same time offer many advantages, which reflect on the welfare of the animals and environment in the region.

The problem of ownership identification of newly born reindeer calves was presented as an example application from the experience of reindeer herding in northern Norway, which lies within the arctic region. It described the process of matching new-born reindeer calves to their mothers, so that they can be given the proper RFID tags of their relevant owners, based on Wi-Fi enabled RFID tags.

This method is expected to reduce the amount of effort and time in performing the identification process under the harsh arctic conditions and paves the way for the implementation of Wi-Fi enabled RFID tags for animal welfare applications.

ACKNOWLEDGEMENT

This research work is undertaken under the Botnia-Atlantica programme, co-funded by the European Regional Development Fund.

REFERENCES

[1] EU. *Botnia- Atlantica Project*. 2013 [retrieved, April 2013]; Available from: http://www.botnia-atlantica.eu/default.asp?lid=1

[2] Karlsson, J., Wark, T., Haake, J., and Björk, P., *The digital zoo: a testbed for wireless video sensor networks*. in *10th Scandinavian Workshop on Wireless Ad-hoc Networks (ADHOC ´11)*. 2011 Johannesberg Castle, Stockholm.

[3] Karlsson, J., *Animal Sensor Network*. 2013 [retrieved, April 2013]; Available from: http://www.tfe.umu.se/english/research/embedded-systems-lab/projects/animal-sensor-networks/.

[4] EU. *Interreg Project*. 2013 [retrieved, April 2013]; http://www.interreg.no/ireg/web.nsf/ShowProject?

[5] Ranjan, P., Saraswat, P., Kumar, A., Polana, S., and Singh, A., *WildCENSE - GPS based animal tracking system*. in *IEEE Xplore 2008*. 2008.

[6] Kwong, K.H., *Practical considerations for wireless sensor networks in cattle monitoring applications.* Computers and Electronics in Agriculture, 2012. **81**: pp. 33-44.

[7] Want, R., *An introduction to RFID technology.* IEEE Pervasive Computing, 2006. **5**(1): pp. 25-33.

[8] Bouet, M.,. *RFID Tags: Positioning Principles and Localization Techniques*. in *Wireless Days, WD '08. 1st IFIP*. 2008. Dubai: IEEE.

[9] Hassan, T. and Chatterjee, S., *A taxonomy for RFID*. in *Proc. of HICSS*. 2006.

[10] Kealy, A., B. Li, Gallagher, T. and A. Dempster (2009). Evaluation of WiFi Technologies for Indoor Positioning Applications. Proceedings of the Surveying & Spatial Sciences Institute Biennial International Conference, Adelaide 2009, pp. 411-421. ISBN: 978-0-9581366-8-6.

[11] Fine, C., Klym, N., Trossen, D., and Tavshikar, M., *The Evolution of RFID Networks: The Potential for Disruptive Innovation* in *http://ebusiness.mit.edu*, M. Sloan School of Management, Editor MIT Sloan, March, 2006.

[12] excITingIP.com, *Location Tracking using Wi-Fi based RFID Tags*, 2012, excITingIP.com.

[13] O'Connor, M.C., *RFID's Urge to Converge.* RFID Journal, 2006. **1** (1).

[14] Lui, T., *Wi-Fi-Based Active RFID Technology & Applications*, 2007, Aeroscout Enterprise visibility solutions.

[15] Baker, N., *ZigBee and Bluetooth - strength and weakness for industrial applications.* Computer control Eng, 2005. **16**: pp. 20-25.

[16] Lu, K. Y., *A plug-and-play data gathering system using ZigBee-based sensor network sensor network.* Computers in Industry, 2011. **62** (7): pp. 719-728.

 ISBN: 978-1-61208-296-7

Fuzzy/PSO Based Washout Filter for Inertial Stimuli Restitution in Flight Simulation

Khaled Fellah, Mohamed Guiatni and Yacine Morsly
Control Laboratory
Ecole Militaire Polytechnique
Bordj El Bahri, Algiers, 16111, Algeria
Email: khaled.fellah@gmail.com, mohamed.guiatni@gmail.com, ymorsly@yahoo.fr

***Abstract*—The aim of this study is to present a new approach using Particle Swarm Optimization (PSO) algorithm and fuzzy logic for motion cueing considering both the physical limits of restitution platforms and realistic sensations. We added the necessary software in order to restitute the specific force based on a virtual aircraft. We used Microsoft Flight Simulator software (MSFS) and built-in data structure and methods. Results using the overall simulation are presented and evaluated on a motion platform. Interesting sensations have been recorded, which enhance the realism of the simulation. The obtained results indicate that the proposed PSO/Fuzzy approach improves the performance of the classical washout filter based motion cueing algorithm.**

Keywords*—*Inertial stimuli; flight simulation; washout filter; fuzzy logic; particle swarm optimization.

I. INTRODUCTION

In the last decades, the use of driving simulation for traffic safety, vehicle design and driver perception studies has expanded rapidly [1] [2] [3]. This is largely because simulation saves engineering time and costs, and can be used for studies of road and traffic safety. Simulation is also a useful and indispensable tool for aviation research and training. It has evolved and matured over the last forty years in equal pace with developments in the aerospace industry. Flight simulation allows pilots to fly in simulated conditions, without the costs and safety issues that go with performing real flight.

In addition, recent psychophysical studies have revealed an unexpectedly important contribution of vestibular cues in distance perception and steering, prompting a re-evaluation of the role of visuo-vestibular interaction in driving simulation studies.

When flight simulation and research are combined, the objective is to measure the human performance in the simulated environment [4]. Research will pose certain requirements on the use of simulation hardware and software. It requires generic tools that can be adjusted to the evolving insight in topics. This implies that flight simulators (hardware) and simulation models (software) used for research will often be a trade-off between realism and flexibility [5]. Therefore, a flight simulator must include an aircraft model, a display capability and control hardware [6]. The aircraft model is implemented as software [7]. Three examples of commonly used flight simulator packages are Microsoft Flight Simulator (sometimes abbreviated to MSFS or FS), FlightGear and XPlane.

The flight controls are hardware providing input to the aircraft model. In most cases they are used along with an aircraft-like control input such as a joystick, yoke and rudder pedals.

While a visual system alone can provide motion cues at low frequency, physical motion stimuli are necessary to provide higher frequency cues in the range sensitive to the vestibular and somatosensory systems.

The adjunction of high fidelity motion cues from a moving platform in conjunction with visual motion cues have been shown to produce a rapid onset of vection, or the illusion of motion, thus reducing the delay associated with visual motion alone.

A key element in providing physical stimuli in flight simulators is the cueing algorithm konwn as washout filter that produces the drive signals used to control the motion system hardware.

In this paper, we propose a new approach for motion cueing considering both the physical limits of restitution platforms and realistic sensations objective. The proposed approach integrates fuzzy logic and Particle Swarm Optimization (PSO) algorithm to the classical washout filter. This approach is an improvement of those presented in the litterature. We added the necessary software in order to restitute the specific force based on a virtual aircraft. We used Microsoft Flight Simulator software (MSFS) and built-in data structure and methods. Results using the overall simulation are presented and evaluated on a motion platform.

This paper is organised as follows: Section II gives a background about the evolution of researches on the washout filter. Section III presents the different frames and a brief overview on the classical washout filter priciple and algorithm. Section IV presents the coordinate transformation of the orientation between the body-fixed simulator reference frame and the inertial reference frame. In Section V, we present the fuzzy logic washout filter design and the optmization algorithm. Section VI presents the Microsoft Flight Simulator software which is used to perform the simulation. Finally, we present the implementation details and some results about the fuzzy logic based motion cueing algorithm for flight simulation.

II. BACKGROUND ON MOTION CUEING ALGORITHMS

Many related researches on washout filter have been presented in the last three decades. Classical washout filter is

 ISBN: 978-1-61208-296-7

the first scheme that has been proposed, which is composed of linear low-pass and high-pass filters. Its advantages are simplicity and easy to adjust. The fixed scheme and parameters of the classical washout filter cause inflexibility of the scheme and make the resulting simulator fail to suit all circumstances. In [8] [9], the authors proposed coordinated adaptive filters, used to coordinate translational and rotational motion.

Nahon and Reid [10] suggested an adaptive washout algorithm with the same scheme as that for classical washout filter and with self-tuning of the filter parameters. Sivan et al. [11] proposed an optimal washout algorithm which takes into account the vestibular system models. This algorithm uses techniques of optimal control and employs a cost function that depends on both sensation error and platform motion. The optimal algorithm designs an optimal structure and a set of optimal parameters subject to the assumptions of human vestibular models and platform limitation by solving the Riccati equation [11]. Telban et al. [12] formulated a linear optimal control problem similar to [11] and solved the Riccati equation in real time so that a scalar coefficient that increases control action can be tuned online. The magnitude of the scalar coefficient depends upon the platform motion.

With large platform motion, the large coefficient increases and results in faster control action. Nehaoua et al. [13] applied classical, adaptive, and optimal algorithms and compared performances of these algorithms in their driving simulator. Authors in [14] propose the use of a linear optimal controller synthesized with a quadratic performance index, which has been applied to a planar model of the Vertical Motion Simulator. Idan and Nahon [15] proposed the use of a robust controller design based on the model of the motion-base dynamics and control in order to compensate for changes or uncertainties in the motion-base dynamics, particularly electrically driven ones with limited bandwidth and power. In [16], a variation of the optimal algorithm was formulated, which incorporate the models of the human vestibular sensation system, i.e., the semicircular canals and otoliths. No matter what kind of platform is used as the simulator, the limited workspace is an important issue in designing the motion cueing algorithm. Several works that increase the efficiency of the platform workspace have been presented.

Huang and Fu [17] proposed a senseless manoeuvre that moves the operator with the acceleration under the threshold value of human perception to conserve the workspace. Liao et al. [18] combined the classical washout filter with an adjustable scaling filter, a yawing washout filter, a dead zone washout filter, and an adaptive washout filter in order to complete the motion planning of the simulator in restricted workspace. A common method to this problem is the fuzzy control-based washout filter proposed by [19] [20], which propose to tune the washout filter parameters based on fuzzy logic in real-time. The cut-off frequencies of the filters are adjusted according to the workspace margins and driving conditions. That algorithm combined characters of motion perception and actual movement of a train to design the fuzzy control rules. However, no optimization process is included in these works. In [21], an optimal washout filter is developed by applying the algorithm of Sivan et al. [11] to the human vestibular system. In this later, the fuzzy controller is used to compensate the filtred signals.

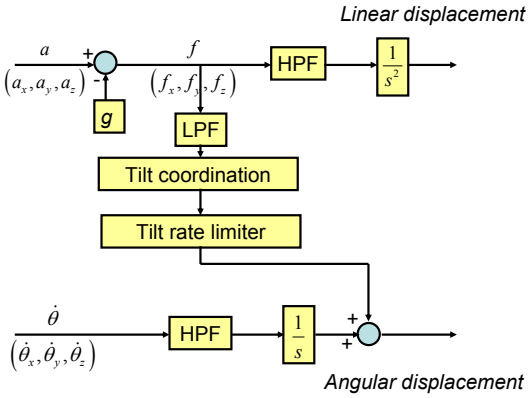


Fig. 1: Classical washout filter flowchart.

The aim of this work is minimize the sensation error produced from the comparison of the human vestibular signals about actual vehicle and simulator. The optimal motion cueing algorithm is then featured by systematic integration of linear filters that are determined through an off-line design process. For the drawback with fixed parameters of the optimal algorithm, the fuzzy control rules are designed to reduce the sensation errors. In order to further eliminate the sensation errors, we propose a new motion cueing algorithm to be used in a research flight simulator using the Particle Swarm Optimization algorithm.

III. CLASSICAL WASHOUT FILTER

A vestibular system, consisting of semicircular canals and otoliths, has a great role in sensing physical motion cues in a motion simulator. It is known that the otoliths are able to detect the high-frequency components of the linearly accelerated motion via a specific force, while the semicircular canals can sense the high-frequency components of the angular motion. Due to the structural limit of the Stewart-Gough platform [22] [8], a motion base has relatively small workspace compared with the operating range of a real aircraft. Because of this limited motion envelope of typical motion bases, filtering is required between the aircraft motion computed from aircraft dynamics and the simulated motion commanded to the motion base. The filter used for this purpose is called a 'washout filter'.

Fig. 1 shows the classical washout algorithm which consists of a High-Pass Filter (HPF) and a Low-Pass Filter (LPF) [23]. The linear acceleration (i.e., specific force) and the angular velocity from the vehicle dynamics are sent to the washout algorithm. The details on how the linear and angular displacements are computed in the washout filter are discussed below. The specific force f_x, f_y, and f_z are computed by subtracting g from the given linear acceleration a. The high-frequency components of the specific force f_x are then obtained by the following 2^{nd}-order HPF:

$$\frac{f_{Hx}}{f_x} = \frac{s^2}{s^2 + 2\zeta_{Hx}\omega_{Hx}s + \omega_{Hx}^2} \quad (1)$$

where ζ and ω represent the damping ratio and the cutoff frequency of the HPF, and the subscript H denotes the high-frequency. Note that f_{Hy} and f_{Hz} in the y and z axes are similarly obtained. This HPF filters out low-frequency motion signals, which tend to lead to large linear displacements that

 ISBN: 978-1-61208-296-7

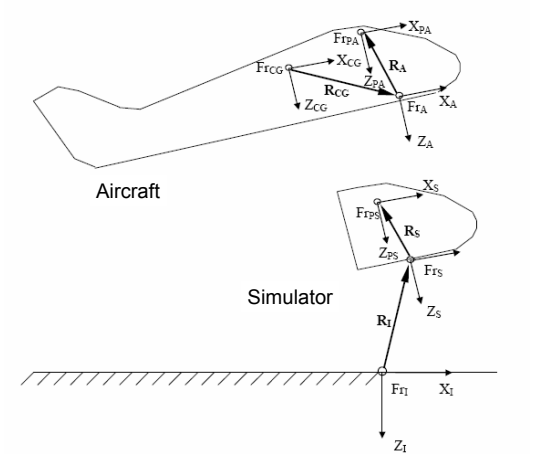


Fig. 2: Frames definition.

cause the motion base to reach its motion limits. The high-frequency motion signals are realized by linear motion of the platform of the motion base by double integration of the filtered signals. Low-frequency components of the specific force in the x axis can be obtained by the following 2nd-order LPF:

$$\frac{f_{Lx}}{f_x} = \frac{\omega_{Lx}^2}{s^2 + 2\zeta_{Lx}\omega_{Lx}s + \omega_{Lx}^2} \tag{2}$$

where the subscript L denotes the low-frequency.

Note that fly is similarly obtained. This low frequency portion is realized by tilting the platform (i.e., angular motion about the $x-$ and $y-$ axes), since a human perceives as if he experienced a specific force when his body tilts. This is referred to as *tilt coordination*. Usually, a pitch angle is set to $\theta_{ytilt} = -f_{Lx}/g$ and a roll angle to $\theta_{xtilt} = -f_{Ly}/g$ for relatively small angles. The rate of tilting is limited to $l - 5°$/s to prevent the semicircular canal from perceiving this signal for tilt coordination as the angular motion cue [10]. In the meantime, the high-frequency components of angular velocity about the $x-$ axis are obtained by the following first-order HPF :

$$\frac{\dot{\theta}_{Hx}}{\dot{\theta}_x} = \frac{s}{s + \omega_{H\theta x}} \tag{3}$$

Note that $\dot{\theta}_{Hx}$ and $\dot{\theta}_{Hx}$ about the y and z axes are similarly obtained. The high-frequency components are integrated to give the angular displacement $(\theta_{Hx}, \theta_{Hy}, \theta_{Hz})$ of the platform. The orientation angle (θ_x, θ_y) about the x and y axes are then obtained by adding the tilt angles $(\theta_{xtilt}, \theta_{ytilt})$ to the angles $(\theta_{Hx}, \theta_{Hy})$ due to angular motion, while θ_z is given as θ_z since no tilt coordination is performed about this axis. It should be noted that the cross feed channel in Fig. 1 includes a rate-limiting element that keeps the tilt rates below 3 °/s.

A. Reference Frames

A series of reference frames are used in the definition of the motion cueing algorithms. These reference frames are adopted from [24], defined below and shown in Fig. 2.

B. Aircraft Center of Gravity

The aircraft center of gravity reference frame Fr_{CG} has its origin at the center of gravity of the aircraft. Frame Fr_CG has an orientation for X_{CG}, Y_{CG}, and Z_{CG} that is parallel to reference frames Fr_S and Fr_A.

C. Simulator Frame

The simulator reference frame Fr_S has its origin at the centroid of the simulator payload platform, i.e., the centroid of the upper bearing attachment points. The origin is fixed with respect to the simulator payload platform. X_S points forward and Z_S points downward with respect to the simulator cockpit, and Y_S points toward the pilot's right hand side. The $x - y$ plane is parallel to the floor of the cockpit.

D. Aircraft Frame

The aircraft reference frame Fr_A has its origin at the same relative cockpit location as the simulator reference frame Fr_S. Fr_A has the same orientation for X_A, Y_A, and Z_A with respect to the cockpit as the simulator frame Fr_S.

E. Inertial Frame

The inertial reference frame Fr_I is earth-fixed with Z_I aligned with the gravity vector g. Its origin is located at the center of the fixed platform motion base. X_I points forward and Y_I points to the right hand side with respect to the simulator pilot.

F. Reference Frame Locations

In Fig. 2, four vectors are illustrated, which define the relative location of the reference frames. R_I defines the location of Fr_S with respect to Fr_I. R_S defines the location of Fr_{PS} with respect to Fr_S. Similarly, R_A defines the location of Fr_{PA} with respect to Fr_A, where $R_A = R_S$. R_{CG} defines the location of Fr_A with respect to Fr_{CG}.

IV. Coordinate Transformations

The orientation between the body-fixed simulator reference frame Fr_S and the inertial reference frame Fr_I can be specified by three Euler angles: $\beta = [\varphi, \theta, \psi]$ that define a sequence of rotations that carry Fr_S into Fr_I. A vector V expressed in the two frames can be related by the transformation matrix L_{SI} (Fr_S to Fr_I), $V^I = L_{SI}V^S$ with:

$$L_{SI} = \begin{bmatrix} c\theta c\psi & s\phi s\theta c\psi - c\phi s\psi & c\phi s\theta c\psi + s\phi s\psi \\ c\theta s\psi & s\phi s\theta s\psi + c\phi c\psi & c\phi s\theta s\psi - s\phi c\psi \\ -s\theta & s\phi c\theta & c\phi c\theta \end{bmatrix} \tag{4}$$

The angular velocity of Fr_S with respect to Fr_I can be related to the Euler angle rates $\dot{\beta}$ by the following expression. Let ω_s^s represent the components of this angular velocity in frame Fr_S, then $\dot{\beta} = T_s\omega_s^s$, where :

$$T_s = \begin{bmatrix} 1 & sin\phi tan\theta & cos\phi tan\theta \\ 0 & cos\phi & -sin\phi \\ 0 & sin\phi/cos\theta & cos\phi/cos\theta \end{bmatrix} \tag{5}$$

The specific force is defined as:

$$f_{pS}^S = a_{pS}^s - g^s \tag{6}$$

 ISBN: 978-1-61208-296-7

where a^s_{pS} is the linear acceleration and g^s is the gravitational acceleration. It is noted that human sense of motion is more sensitive to high-frequency components of the motion than low-frequency ones.

A. Nonlinear Input Scaling

Limiting and scaling are applied to both aircraft translational input signals a^s_s and rotational input signals ω^s_s. Limiting and scaling modify the amplitude of the input uniformly across all frequencies. Limiting is a nonlinear process that clips the signal so that it is limited to be less than a given magnitude. Limiting and scaling can be used to reduce the motion response of a flight simulator. A third-order polynomial scaling has been implemented in the simulator motion cueing algorithms.

When the magnitude of the input to the simulator motion system is small, the gain is desired to be relatively high, or the output will be below the pilot's perception threshold. When the magnitude of input is high, the gain is desired to be relatively low or the simulator may attempt to go beyond the hardware limits. Let us define the input as x and the output as y. Now, define x_{max} as the expected maximum input and y_{max} as the maximum output, and s_0 and s_1 as the slopes at $x = 0$ and $x = x_{max}$, respectively. Four desired characteristics for the nonlinear scaling are expressed as:

$$\begin{cases} x = 0 \Rightarrow y = 0 \\ x = x_{\max} \Rightarrow y = y_{\max} \\ \dot{y}|_{x=0} = s_0 \\ \dot{y}|_{x=x_{\max}} = s_1 \end{cases} \tag{7}$$

A third-order polynomial is then employed to provide functions with all the desired characteristics. This polynomial will be of the form:

$$y = c_0 + c_1 x + c_2 x^2 + c_3 x^3 \tag{8}$$

where :

$$\begin{cases} c_0 = 0 \\ c_1 = s_0 \\ c_2 = x^{-2}_{\max}\left(3y_{\max} - 2s_0 x_{\max} - s_1 x_{\max}\right) \\ c_3 = x^{-3}_{\max}\left(s_0 y_{\max} - 2y_{\max} + s_1 x_{\max}\right) \end{cases} \tag{9}$$

V. Proposed Washout Filter Design

Washout filters have parameters that can be adjusted to alter the motion responses of a simulator. Since the parameters of most classical washout filters remain fixed during their operation, they cannot cope with various flight conditions efficiently. On-line tuning of the washout filters can enable the simulator to offer better simulator motions within the limited motion range. As part of the design of the experiment, the underlying factor for the entire project was the sensed specific force that results from the aircraft motion. This section focuses on design of a fuzzy model, which shall generate the adequate parameters for the eight filters.

A. Fuzzy Logic Washout Filter Design

1) Fuzzification: In this research, fuzzy logic [25] is suggested to tune the filter parameters, since it is known that the fuzzy logic is effectively applicable to a system in which the input-output relations are not clearly identified. To this end, the fuzzy logic is used which takes the Displacement Limit (DL), the Angle Limit (AL) and the Low-Frequency Specific Force (LFSF) as inputs and then provides the proper parameters of the six HPFs and the two LPFs as outputs. Fig. 3 illustrates the flowchart showing this fuzzy logic system. Triangular functions are used for each variable. There were 5 linguistic terms for each input variable. These terms are:

- Displacement limit (DL) & angle limit (AL): VF (Very Far), F(Far), M(Medium), N(Near) and (VN) Very Near.
- Low-frequency specific force (LFSF): VS (Very Small), S (Small), M (Medium), B (Big) and VB (Very Big).

2) Rule-based system: Once the proper inputs were created, the next step is to create the output's characteristic behavior. The output variables was classified as :

- HPF cutt-off frequencies: VS (Very Small), S (Small), M (Medium), B (Big) and VB (Very Big).
- LPF cutt-off frequencies: VSF (Very Small with Far Limit), VSM (Very Small with Medium Limit), VSN (Very Small with Near Limit), SF (Small with Far Limit), SM (Small with Medium Limit), SN (Small with Near Limit), MF (Medium with Far Limit), MM (Medium with Medium Limit), MN (Medium with Near Limit), BF (Big with Far Limit), BM (Big with Medium Limit), BN (Big with Near Limit), VBF (Very Big with Far Limit), VBM (Very Big with Medium Limit), VBN (Very Big with Near Limit).

The next step is to create a rule base that would govern the operation of the fuzzy system. The proper conditions must be created to implement a system that will allow for perceptible specific force reproduction while taking into account the motion limits of restitution platform. The fuzzy rules are shown in Tables I and II.

In order to select the optimal fuzzy membership parameters, we use the Particle Swarm Optimization algorithm (PSO). Some of the attractive features of the PSO include the ease of

TABLE I. Rule based system for the HPF cutt-off frequencies selection.

Displacement limit (DL)	VF	F	M	N	VN
or angle limit (AL)	VF	F	M	N	VN
HPF cutt-off frequencies	VS	S	M	B	VB

TABLE II. Rule based system for the LPF cutt-off frequencies selection.

Angle limit (AL) Low-frequency specific force (LFSF)	VF	F	M	N	VN
VS	VBN	VBF	BN	BM	BF
S	MN	MM	MF	SN	SM
M	VBF	BN	BM	BF	SF
B	MM	MF	SN	SM	VSN
VB	BN	BM	BF	VBM	VBF

 ISBN: 978-1-61208-296-7

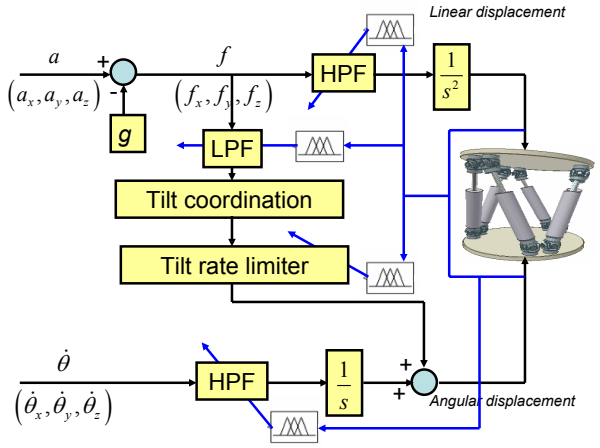


Fig. 3: Fuzzy/washout filter architecture.

implementation and the fact that no gradient information is required. In addition, PSO has the same effectiveness (finding the true global optimal solution) as the Genetic Algorithms (GA) but with significantly better computational efficiency (less function evaluations).

B. Particle Swarm Optimization Algorithm (PSO)

PSO is a stochastic optimization algorithm [26] [27] [28]. The main idea of the PSO is the mathematical modeling and simulation of the food searching activities of a flock of birds. In the multidimensional space, each particle is moved toward the optimal point by changing its position according to a velocity. The velocity of a particle is calculated by three components: inertia, cognitive, and social. The inertial component simulates the inertial performance of the bird to fly in the previous direction. The cognitive component models the memory of the bird about its previous best position. The social component models the memory of the bird about the best position among the particles. The particles move around the multidimensional search space until they find the optimal solution. Based on the above discussion, the mathematical model for PSO is given as follows:

$$\begin{cases} V_i^{t+1} = \omega . V_i^{t+1} + c_1 . rand_1\,(.)\,.\,(Pbest_i - X_i^t) \\ \qquad\qquad +c_2 . rand_2\,(.)\,.\,(Gbest - X_i^t) \\ X_i^{t+1} = X_i^t + V_i^{t+1}, \quad i = 1, 2, 3, ..., N_{swarm} \end{cases} \tag{10}$$

where, i is the index of each particle, t is the current iteration number, $rand_1$ and $rand_2$ are random numbers between 0 and 1. $Pbest_i$ is the best previous experience of the i^{th} particle that is recorded. $Gbest$ is the best particle among the entire population. N_{Swarm} is the number of the swarms. Constants c_1 and c_2 are the weighting factors of the stochastic acceleration terms, which pull each particle towards the $Pbest_i$ and $Gbest$. ω is the inertia weight. As indicated in (10), there are three tuning parameters; ω, $c1$, and c_2 that each of them has a great impact on the algorithm performance. The inertia weight ω controls the exploration properties of the algorithm. The learning factors c_1 and c_2 determine the impact of the personal best $Pbest_i$ and the global best $Gbest$, respectively.

If $c_1 > c_2$, the particle has the tendency to converge to the best position found by itself $Pbest_i$ rather than the best position found by the population $Gbest$, and vice versa. Most implementations use a setting with $c_1 = c_2 = 2$. To implement the PSO algorithm to solve our problem, the following steps were taken:

- *Step 1:* The initial population and initial velocity for each particle should be generated randomly.
- *Step 2:* The objective function is to be evaluated for each individual.
- *Step 3:* The individual that has the minimum objective function should be selected as the global position.
- *Step 4:* The i^th individual is selected.
- *Step 5:* The best local position ($Pbest$) is selected for the i^th individual.
- *Step 6:* The modified velocity for the i^th individual needs to be calculated based on the local and global positions and (10)).
- *Step 7:* The modified position for the i^th individual should be calculated based on (10) and then checked with its limit.
- *Step 8:* If all individuals are selected, go to the next step, otherwise $i = i + 1$ and go to step 4.
- *Step 9:* If the current iteration number reaches the predetermined maximum iteration number, the search procedure is stopped, otherwise go to step 2.

The objective function is to minimize the sensation error between the spectific force estimated on the aircraft and the one restituted on the simulator based on the proposed fuzzy washout algorithm with respect to the following constraints:

- The negative acceleration is limited to $0.17m/s^2$ after high pass filtering.
- The washout rate is limited to $0.048m/s$.
- The output of the low pass filter after tilt coordination is limited to $5°/s$.

The last $Gbest$ is the solution of the problem.

VI. Experiments and Implementation Details

A. Virtual aircraft model and instrumentation

The use of non-flight-certified Commercial-Off-The-Shelf (COTS) solutions has proved its fidelity and coordination characters as a flight training device [29]. MSFS is a flight simulator program, marketed and often seen as a video game. However, it is less a game than an immersive virtual environment since it is very realistic, (see Fig. 4). Its first version appeared in 1982, whereas, its most recent versions, Century of Flight and Flight Simulator X appear respectively in 2003 and 2006 [30].

The long history, the consistent popularity and the open nature of flight simulator structure have encouraged a very large body of freeware and payware add-on packages to be developed. These add-ons, widely available over the internet, are very helpful because they, not only, permit to change internal aspects of the simulator (airplanes, scenery...), but also to interface it with external software and hardware such as

 ISBN: 978-1-61208-296-7

Fig. 4: Virtual Mi24 helicopter under the MSFS environment.

Fig. 5: The *MOOG 6DOF 2000E 170 E131* motion platform.

homebuilt cockpits. Among the most famous add-ons, we cite a dynamic link library add-on made by Pete Dowson [31], called Flight Simulator Universal Inter-Process Communication (FSUIPC). This module is designed to allow external (i.e., separate) programs to communicate with and control MSFS in real time. In other words, they permit to read from and write in MSFS while it's running and place it in a 64 Kb buffer. So to achieve a variable we have only to know its offset (address) in this buffer. However, before writing or reading data from FSUIPC, we have to scale it to the desirable unit.

Our algorithm was developed to achieve the desired motion cues at an update rate of 60 Hz. Since the computer image generator, which provides the out-the window visual imagery to the simulator pilot, also runs at 60 Hz, the motion cues would be synchronous with the visual cues. The HPF and PHF filters are implemented using the tustin operator which consists to replace the Laplace operator by $s = \frac{2}{Ts}\frac{1-z^{-1}}{1+z^{-1}}$, where z is the discret operator, and Ts is the sample time. The proposed algorithm was successfully implemented on the *MOOG 6DOF 2000E 170 E131* motion platform (see Fig.5).

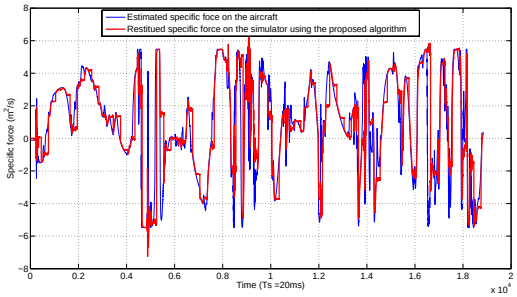


Fig. 6: Real specific force *vs* restituted specific force using the proposed algorithm.

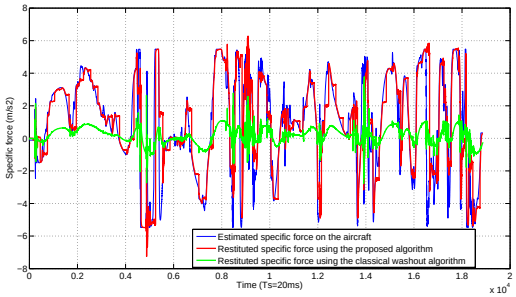


Fig. 7: Restituted specific force using the classical washout algorithm *vs* restituted specific force using the proposed algorithm.

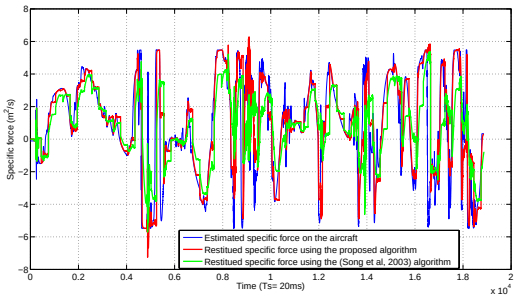


Fig. 8: Restituted specific force using the [19] algorithm *vs* restituted specific force using the proposed algorithm.

B. Results and Discussion

Figs. 6 to 9 show an example of simulated manoeuvres for a MI24 helicopter. Fig. 6 presents comparison between the specific force estimated on the aircraft and that restituted on the simulator based on the proposed algorithm. The specific force is estimated using the vestibular model as explained above.

Fig. 7 shows a comparison between the produced specific force using the classical washout algorithm and that produced using the proposed algorithm. We remark from this figure that the specific force signal produced using our algorithm is highly important and closer to the real signal in comparison with the result obtained using the classical algorithm.

While comparing the obtained results in Fig. 8, we can see that the proposed algorithm allows to get a specific force closer to the one estimated on the aircraft in comparison with the results obtained using the approach proposed in [19].

 ISBN: 978-1-61208-296-7

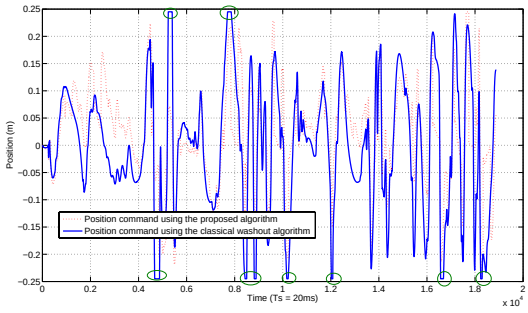


Fig. 9: Comparison of the position cammands using the classical washout algorithm *vs* restituted the position cammands using the proposed algorithm.

Therefore, it allows to the pilot to get a realistic sensation on the simulator. In addition, the magnitude of position commands sent to the motion platform are always within its displacement limits, as seen from Fig. 9. In this way, we can preserve the motion platform from damage and maintain its functionality.

VII. CONCLUSION

This paper reported a successful achievement of a new fuzzy washout algorithm implementation for inertial stimuli restitution in Flight Simulation. The integration of the PSO algorithm has allowed to select the optimal parameters of the fuzzy models. The related built-in software have been integrated successfully with COTS software that provides the aircraft dynamics in a virtual environment. The different simulations proved the effectiveness of the proposed algorithm.

REFERENCES

[1] A. Kemeny and F. Panerai, "Evaluating perception in driving simulation experiments," in *Trends in Cognitive Sciences*, vol. 7, no. 1, 2003, pp. 31–37.

[2] D. Katzourakis, J. de Winter, S. de Groot, and R. Happee, "Driving simulator parameterization using double-lane change steering metrics as recorded on five modern cars," in *Simulation Modeling Practice and Theory*, vol. 26, 2012, pp. 96–112.

[3] M. Hoedemaeker and K. Brookhuis, "Behavioural adaptation to driving with an adaptive cruise control (acc)," in *Transportation Research Part F*, vol. 1, no. 2, 1998, pp. 95–106.

[4] S. Valenti Clari, R. C. Ruigrok, B. W. Heesbeen, and J. Groeneweg, "Research flight simulation of future autonomous aircraft operations," in *Proceedings of the Conference of European Aerospace Societies (CEAS)*, vol. 2, 8-11 Dec 2002, pp. 1126–1134.

[5] J. Hoekstra, "The smart software-simple hardware concept for maximum flexibility in research flight simulation," *Proceedings of the Conference of European Aerospace Societies (CEAS)*, vol. NLR-TP-96215, p. 16, Amsterdam, the Netherlands 1995.

[6] H. Teufel, H. Nusseck, K. Beykirch, J. Butler, M. Kerger, and H. Blthoff, "MPI Motion Simulator: Development and Analysis of a Novel Motion Simulator," in *AIAA Modeling and Simulation Technologies Conference and Exhibit*, Hilton Head, South Carolina, 20-23 August 2007, pp. 1–11.

[7] F. Holzapfel, I. Sturhan, , and G. Sachs, "Low-cost pc based flight simulator for education and research," *AIAA Modeling and Simulation Technologies Conference and Exhibit*, no. 1, p. 1540, 5-8 August 2002.

[8] R. Parrish, J. Dieudonne, D. Martin, and R. Bowles, "Coordinated adaptive filters for motion simulators," in *Summer Computer Simulation Conference*, 1973, pp. 295–300.

[9] R. Parrish, J. Dieudonne, and R. Bowles, "Coordinated adaptive washout for motion simulators," in *J. Aircraft*, vol. 12, no. 1, 1975, pp. 44–50.

[10] M. A. Nahon, L. D. Reid, and J. Kirdeikis, "Adaptive Simulator Motion Software with Supervisory Control," in *Journal of Guidance, Control and Dynamics*, vol. 15, no. 2, 1992, pp. 376–383.

[11] R. Sivan, J. Ish-Shalom, and J. Huang, "An optimal control approach to the design of moving flight simulators," in *IEEE Trans. On Systems, Man, and Cybernetics*, vol. 12, no. 6, 1982, pp. 818 – 827.

[12] R. Telban, F. Cardullo, and J. Houck, "A nonlinear, human-centered approach to motion cueing with a neurocomputing solver," in *AIAA Modeling and Simulation Technologies Conference and Exhibit*, Monterey, California, 2002.

[13] L. Nehaoua, H. Mohellebi, A. Amouri, H. Arioui, S. Espié, and A. Kheddar, "Design and control of a small-clearance driving simulator," in *IEEE Transactions on Vehicular Technology*, vol. 57, no. 2, 2008, pp. 736 – 746.

[14] W. Sturgeon, "Controllers for aircraft motion simulators," in *Journal of Guidance and Control*, vol. 4, no. 2, 1981, pp. 184–191.

[15] M. Idan and M. Nahon, "Off-line comparison of classical and robust flight simulator motion control," in *J. Guidance, Control and Dynamics*, vol. 22, no. 5, 1999, pp. 184–191.

[16] R. Telban, F. Cardullo, and J. Houck, "Developments in human centered cueing algorithms for control of flight simulator motion systems," in *Modeling and Simulation Technologies Conference and Exhibit*, 1999, pp. 1–10.

[17] C. Huang and L. Fu, "Human vestibular based (HVB) senseless maneuver optimal washout filter design for VR-based motion simulator," in *IEEE International Conference on Systems, Man and Cybernetics*, 2006, pp. 4451–4458.

[18] C. Liao, C. Huang, and W. Chieng, "A novel washout filter design for a six degree-of-freedom motion simulator," in *JSME International Journal, Series C*, vol. 47, no. 2, 2004, pp. 626 – 636.

[19] J. Song, U. Jung, and H. Ko, "Washout algorithm with fuzzy-based tuning for a motion simulator," in *KSME International Journal*, vol. 17, no. 2, 2003, pp. 221–229.

[20] X. Wang, L. Li, and W. Zhang, "Research on fuzzy control washout algorithm of locomotive driving simulator," in *Proceedings of the 7th World Congress on Intelligent Control and Automation*, Chongqing, 2008, pp. 3737 – 3741.

[21] S. Chen and L. Fu, "An optimal washout filter design with fuzzy compensation for a motion platform," in *18th IFAC World Congress*, Milano, Italy, 2011, pp. 8433–8438.

[22] D. Stewart, "A platform with six-degrees-of-freedom," in *Proc. Inst. of Mechanical Engineers, part 1*, vol. 180, no. 5, 1966, pp. 371–386.

[23] P. Grant and L. Reid, "Motion washout filter tuning: Rules and requirements," in *AIAA Flight Simulation Technologies*, vol. 34, no. 2, 1997, pp. 145–151.

[24] W. Wu, "Development of cueing algorithm for the control of simulator motion systems," Ph.D. dissertation, State University of New York at Binghamton, Binghamton, NY, 1997.

[25] R. Precup and H. Hellendoorn, "A survey on industrial applications of fuzzy control," in *Computers in Industry*, vol. 62, no. 3, 2011, pp. 213–226.

[26] J. Kennedy and R. Eberhart, "Particle swarm optimization," in *IEEE Int Conf Neural Networks*, 1995, pp. 1942 – 1948.

[27] P. Bajpai and S. Singh, "Fuzzy adaptive particle swarm optimization for bidding strategy in uniform price spot market," in *IEEE Trans Power Syst*, vol. 22, no. 4, 2007, pp. 2152–2160.

[28] T. Niknam, "A new fuzzy adaptive hybrid particle swarm optimization algorithm for non-linear, non-smooth and non-convex economic dispatch problem," in *Journal of Applied Energy*, vol. 87, 2010, pp. 327–339.

[29] S. Zheng, S. Zheng, and J. Han, "COTS and design pattern based high fidelity flight simulator prototype system," in *Journal of Computers*, vol. 6, no. 1, January 2011, pp. 28–35.

[30] "http://www.microsoft.com/games/flightsimulator/," *Microsoft Flight Simulator 2004*, [retrieved: July, 2013].

[31] "http://www.schiratti.com/dowson," [retrieved: July, 2013].

 ISBN: 978-1-61208-296-7

Real-Time Underwater Communication Technique for Energy Efficient Ocean Monitoring

Ranjitha Parameshwaraiah, Ramya Ramesh and Narendra Kumar G
Dept. of Electronics & Communication,
University Visvesvaraya College of Engineering, Bangalore University.
Bangalore, India.
Email: ranjitha040391@gmail.com, ramya161091@gmail.com, gnarenk@yahoo.com

***Abstract*—The need for underwater communications exists in the areas of pollution monitoring, ocean sampling networks, tactical surveillance applications, exploration of natural undersea resources, mine reconnaissance and predicting wave tides. Deployment of underwater sensor networks for real-time investigation is the major challenge. There is a need to deploy underwater networks that will enable real-time monitoring of selected ocean areas, remote configuration and interaction with onshore human operators. This can be obtained by connecting underwater instruments by means of wireless links based on acoustic communication is the most versatile physical link for continuous wireless sensor networks in underwater scenarios. In this paper large-scale underwater Sensor Networks (UWSN) and Underwater Ad-hoc Networks (UANETs) using Solar-Powered Autonomous Underwater Vehicles (SAUV) to explore the oceanic environment is proposed. Kong Wobbler base station with acoustic communication devices is considered, which locates the pre-deployed underwater sensor modules through acoustic communication. The sensor modules are installed with various acoustic sensors and video capturing devices to study the underwater resources as well as for surveillance needs for predicting the environmental conditions highlighting potential applications of seismic monitoring that provides a better disaster warning mechanism. The simulation results are encouraging as this approach is extremely helpful in surveillance; intruders are tracked and real-time data streaming is processed.**

Keywords - Underwater Ad-hoc Networks (UANET's); Underwater Sensor Networks (UWSN); Solar-Powered Autonomous Underwater Vehicles (SAUV); Acoustic Communication; Underwater Acoustic Sensor Networks (UW-ASN); Geographic Adaptive Fidelity (GAF) Protocol; Kong Wobbler.

I. INTRODUCTION

Sensor networks have the promise of revolutionizing many areas of science, industry, and government with their ability to bring computation and sensing into the physical world. The ability to have small devices distributed near the objects being sensed brings new opportunities to observe. Monitoring the underwater environment is vital in predicting wave tides, pollution monitoring, oceanic data collection, tactical surveillance applications, disaster prevention and exploring natural resources. While sensor networks are beginning to be fielded in applications today on the ground, underwater operations remain quite limited by comparison. The largely unexplored vastness of the ocean, covering about 79% surface of the earth, has fascinated human race for a very long time. The traditional approach for ocean-bottom or ocean-column monitoring is to deploy underwater sensors that record data during the monitoring mission and recover the instruments finds disadvantages:

- Real-time monitoring is critical especially in surveillance or in environmental monitoring applications such as seismic monitoring wherein the recorded data cannot be accessed until the instruments are recovered, which may happen several months after the beginning of the monitoring program.
- No interaction is possible between onshore control systems and the monitoring instruments which impedes adaptive tuning of the instruments nor it is possible to reconfigure the system.
- In case of failures, it is not possible to detect them before the instruments are recovered which leads to the complete failure of a monitoring mission.
- The amount of information that can be recorded during the monitoring mission by every sensor is limited by the capacity of the on-board storage devices in the instrument [1].

The paper is organized as follows. Section II illustrates the Kong Wobbler's design, which carries the base station for information relay. Section III and Section IV explain the internal architecture of the underwater sensors and SAUV. Section V presents the characteristics of the suggested acoustic communication channel for the underwater scenario; to assess the aqueous environment; its role and function in the need for large-scale, long term and distributed information collection networks for periodic oceanic monitoring. The large scale aquatic applications demand us to build UWSN and UANETs to explore the inhibited oceans. Energy efficient routing protocols are the most important criteria for the design of underwater sensor networks, since the sensor nodes will be powered by batteries with limited power capacity. Power failure of a sensor node not only affects the node itself but also its ability to forward data packets to the other nodes. Hence Section VI is dedicated to Energy Efficient, Geographical Adaptive Fidelity (GAF) protocol [3]. A general layout is presented in the Section VII. Sections VIII and IX, explain the Implementation of the work and numerical results obtained by Simulation in the Network Simulator (NS2) [2], respectively. Conclusions are summarized in Section X.

 ISBN: 978-1-61208-296-7

II. KONG WOBBLER

A wireless access point with Kong Wobbler structure is made to float on the surface of sea around the area of investigation to which the structure of the base station is fixed on the top. The weight of the base station is less than the Wobbler base structure, maintaining a ratio of 1:4 to prevent the whole structure from toppling. Kong Wobbler is made up of non toxic Food and Drug Administration (FDA) approved polypropylene chosen for its overall strength, impact absorption, sound deadening and non toxicity on the surface of sea made up of high strength polymer and is Bisphenol A (BPA) and phthalate free. Kong Wobbler is filled with compressed CO_2 that allows the structure to float and uses play grade sand for its weight. The sand remains in a compartment that has been permanently sealed using ultrasonic technology. The Kong Wobbler is made to float on the surface of sea and anchored to the sea bed through cables to keep in position such that due to its unique structure it floats vertically. They are hand-launched over the side of a ship or air dropped in the area of investigation. The column of fluid has greater pressure at the bottom of the column of ocean than at the top. This difference in pressure results in a net force that tends to accelerate the Kong Wobbler structure upwards. The magnitude of that force is equal to the difference in the pressure between the top and the bottom of the column, and is also equivalent to the weight of the fluid that would otherwise occupy the column. For this reason, if the density of the structure is greater than that of the fluid in which it is submerged, tends to sink. The buoyancy of the Kong Wobbler exceeds its weight and tends to rise. Density is maintained lesser than the liquid and shaped appropriately so that force can keep the whole structure afloat. The floating Kong Wobbler tends to restore itself to an equilibrium position after a small displacement. It has vertical stability, in case it is pushed down slightly, which will create a greater buoyancy force and unbalanced by the weight force, will push the object back up. Rotational stability is of great importance as given a small angular displacement, the structure returns to its original position [3].

Kong Wobbler is of less density than the seawater and will float upward until it reaches the surface of the seawater at which position, only part of it is submerged. Wobbler is made out of such material, displaces a lot of water on the surface of ocean as the force of the water trying to get into the space keeps the wobbler afloat.

Submerged fraction depends on the density of the Kong Wobbler, as compared to the seawater. Hence:

$$\rho_s V_s = \rho_o V_o (Archimede's principle) \tag{1}$$

where ρ_s the density of the seawater, V_s the volume submerged, ρ_o the density of the Kong Wobbler, V_o the volume of the whole structure,
The Submerged volume is maintained around 0.5 times the total volume of the whole structure, as the density of the whole Kong Wobbler lies in between the density of polypropylene and compressed CO_2.

ρ_s = 1027kg/m3, Density of seawater at the surface.
ρ_o= 946kg/m3, Density of polypropylene.
ρ_{co_2}= 1.52 kg/m3, Density of compressed CO_2
As the Wobbler is displaced vertically, a parcel of fluid in contact is also displaced. The buoyancy force F acting on the Kong Wobbler is the difference between the weight of the displaced parcel of fluid with density ρ', (gV ρ') and the weight of the surrounding water with density ρ_2, (gV ρ_2), where V is the volume of the parcel.

$$F = gV(\rho_2 - \rho') \tag{2}$$

The acceleration of the displaced parcel is:

$$a = \frac{F}{m} = g\frac{(\rho_2 - \rho')}{\rho'} \tag{3}$$

but

$$\rho_2 = \rho + \left(\frac{d\rho}{dz}\right)_{water} \delta z \tag{4}$$

$$\rho' = \rho + \left(\frac{d\rho}{dz}\right)_{parcel} \delta z \tag{5}$$

Using (4) and (5) in (3), ignoring terms proportional to δz^2,

$$E = \frac{-1}{d\rho}\left[\left(\frac{d\rho}{dz}_{water}\right) - \left(\frac{d\rho}{dz}_{parcel}\right)\right] \tag{6}$$

where $E = \frac{-a}{(gdz)}$ is the stability of the water column. This can be written in terms of the measured temperature and salinity t(z), S(z) in the water column [3].

$$E = \alpha\left(\frac{dt}{dz} - g\rho\Gamma\right) - \beta\frac{dS}{dz} \tag{7}$$

where

$$\alpha = \frac{-1}{\rho}\left(\frac{\partial\rho}{\partial t}\right) \mid_{S,p}$$

$$\beta = \frac{-1}{\rho}\left(\frac{\partial\rho}{\partial S}\right) \mid_{t,p}$$

$$\Gamma = \left(\frac{\partial t}{\partial p}\right) \mid_{adiabatic}$$

with α, the thermal expansion coefficient, β the saline contraction coefficient, and Γ is the adiabatic lapse rate, the change of temperature with pressure as the water parcel moves without exchanging heat with it's surroundings. p is pressure, t is temperature in celsius, ρ is density, and S is salinity. Few kilometers from the surface of the ocean, stability is large such that the first term in (6) is much larger than the second. The first term is proportional to the rate of change of density of the water column; the second term is proportional to

 ISBN: 978-1-61208-296-7

the compressibility of sea water, which is very small. Hence, neglecting the second term, stability equation reduces to:

$$E \approx \frac{-1}{\rho}\frac{d\rho}{dz} \quad (8)$$

The approximation used to derive (8) is valid for

$$E > 50 * 10^{-}8/m$$

Below, about a kilometer in the ocean, the change in density with depth is so small that we must consider the small change in density of the parcel due to changes in pressure as it is moved vertically, and (7) must be used [4]. Stability is defined such that

$$E > 0, Stable$$

$$E = 0, Neutralstability$$

$$E < 0, Unstable$$

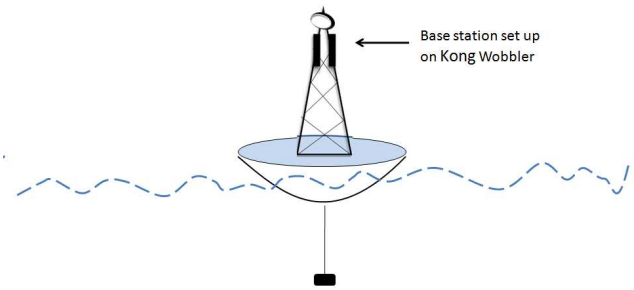


Fig. 1. Kong Wobbler structure

III. SENSORS

Sensor nodes collect data through their sensors. One possible approach to node deployment and communicate with other nodes through short-range acoustic couplers. They have batteries, but for long-term operation they spend most of their life asleep. All nodes are integrated with temperature, vibration, pressure, viscosity, turbidity, seismic, proximity, chemical & gas, fluid flow, speed, water level, altitude and visibility sensors to measure the different parameters in the marine environment which are small, robust, inexpensive, low power consuming yet efficient. Real-time readings are taken from all the sensors to check readings crossing the threshold values to take necessary actions. The typical internal architecture of an underwater sensor is shown in Fig. 2. It consists of a main controller/CPU (Central Processing Unit), which is interfaced with an oceanographic instrument or sensor through a sensor interface circuitry. The controller receives data from the sensor and it can store it in the onboard memory, process it, and send it to other network devices by controlling the acoustic modem. The electronics are usually mounted on a frame which is protected by a PVC housing. Sensor components are protected

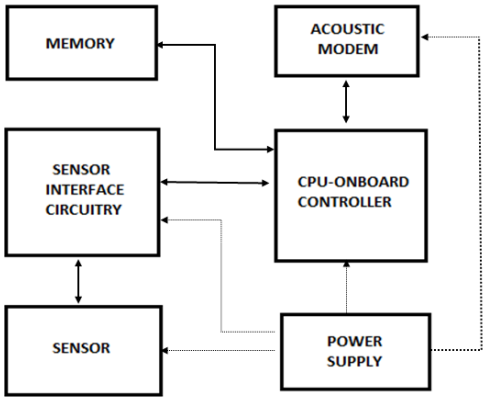


Fig. 2. Sensor node Architecture

by bottom-mounted instrument frames that are designed to permit azimuthally omnidirectional acoustic communications, and protect sensors and modems from potential impact of trawling gear, especially in areas subjected to fishing activities. The protecting frame is designed so as to deflect trawling gear on impact, by housing all components beneath a low-profile pyramidal frame. Underwater sensors include sensors to measure the quality of water and to study its characteristics such as temperature, density, salinity (interferometric and refractometric sensors), acidity, chemicals, conductivity, pH (magnetoelastic sensors), oxygen (Clark-type electrode), hydrogen, dissolved methane gas (METS), and turbidity. Disposable sensors exist that detect ricin, the highly poisonous protein found in castor beans and thought to be a potential terrorism agent. DNA microarrays can be used to monitor both abundance and activity level variations among natural microbial populations. Other existing underwater sensors include hydrothermal sulfide, silicate, voltammetric sensors for spectrophotometry, gold-amalgam electrode sensors for sediment measurements of metal ions (ion-selective analysis), amperometric microsensors for H_2S measurements for studies of anoxygenic photosynthesis, sulfide oxidation, and sulfate reduction of sediments. In addition, force/torque sensors for underwater applications requiring simultaneous measurements of several forces and moments have also been developed, as well as quantum sensors to measure light radiation and sensors for measurements of harmful algal blooms [5].

IV. SOLAR-POWERED AUTONOMOUS UNDERWATER VEHICLES (SAUV)

The SAUV [6] is a solar powered AUV designed for long endurance missions that require monitoring, surveillance/station keeping, with real-time bi-directional communications to shore.

- The SAUV is a solar-powered autonomous vehicle which is equipped with rechargeable lithium ion batteries to allow maximum mission endurance even under conditions where minimal solar radiation is available.

 ISBN: 978-1-61208-296-7

- Operate autonomously at sea for extended periods of time from weeks to months. Typical missions require operation at night and solar energy charging of batteries during daytime.
- In case of failure of Kong Wobbler, SAUVs Communicate with a remote operator on daily basis via the Satellite phone or the RF radio or the acoustic telemetry.
- Operate at speed up to about 3 knots when needed and cruise at speed of about 1 knot.
- Battery system is to provide a total capacity of around 1.5 kwh.
- Capability to acquire GPS updates on the ocean surface and compute SAUV position at all times using GPS.
- Capability to maintain fixed depth and fixed altitude and to smoothly vary depth or altitude profile.
- Capability to log and upload all sensor data correlated in time and SAUV geodetic position.
- Provide sufficient volume, power, interfaces and software hooks for future payload sensors.
- Allow user to program missions easily using a Laptop PC and provide for graphical display of mission [6].

Fig. 3. SAUV

V. Acoustic Communication

Wireless underwater communication is challenging task with the growing need for underwater surveillance and develop persistent long-term ocean observation has led to many underwater wireless technologies. Present underwater communication involves transmission of data in the form of optical waves, electromagnetic or sound waves. Optical waves involved in underwater communication are generally limited to very short ranges because of the strong backscatter from the suspended particles in the ocean, severe absorption by water at optical frequencies and high level of ambient light in the upper part of the water column. Even the clearest water has 1000 times the attenuation of clear air and turbid water has more than 100 times the attenuation of densest fog. Electromagnetic waves in radio frequencies, does not work underwater due to high conducting nature of the medium, especially sea water. Average conductivity of sea water is 4 mhos/meter, therefore the attenuation for 2.4GHz is around 1695dB/meter which is not feasible. Underwater acoustic communications is an important alternative to Radio-Frequency (RF) communications. It is the most versatile and widely used technique in underwater wireless communication which has low attenuation of sound in water used as the primary carrier for underwater wireless communication systems that holds well in thermally stable and deep water settings [3]. It is seen that in simple acoustic propagation models that multi-hop routing saves energy in underwater networks with respect to single hop communications, especially with distances of the order of some kilometers.

The challenges posed by the underwater channels for underwater sensor networking are :

1) Path loss
   - Attenuation: Is mainly provoked by absorption due to conversion of acoustic energy into heat, which increases with distance and frequency. It is also caused by scattering and reverberation (on rough ocean surface and bottom), refraction, and dispersion (due to the displacement of the reection point caused by wind on the surface). Water depth plays a key role in determining the attenuation.
   - Geometric Spreading: This refers to the spreading of sound energy as a result of the expansion of the wave fronts. It increases with the propagation distance and is independent of frequency. There are two common kinds of geometric spreading: spherical (omni-directional point source), and cylindrical (horizontal radiation only).
2) Noise
   - Man made noise. This is mainly caused by machinery noise (pumps, reduction gears, power plants, etc.), and shipping activity (hull fouling, animal life on hull, cavitation).
   - Ambient Noise. Is related to hydrodynamics (movement of water including tides, currents, storms, wind, rain, etc.), seismic and biological phenomena.
3) Multi-path
   - Multi-path propagation may be responsible for severe degradation of the acoustic communication signal, since it generates ISI (Inter-Symbol Interference).
   - The multi-path geometry depends on the link conguration. Vertical channels are characterized by little time dispersion, whereas horizontal channels may have extremely long multi-path spreads, whose value depend on the water depth.
4) High delay and delay variance:
   - The propagation speed in the UW-A channel is order of five, magnitude lower than the radio channel. This large propagation delay (0.67 s=km) can reduce the throughput of the system considerably.

 ISBN: 978-1-61208-296-7

- The very high delay variance is even more harmful for efficient protocol design, as it prevents from accurately estimating the Round Trip Time (RTT), key measure for many common communication protocols.

5) Doppler spread
   - The Doppler frequency spread can be signicant in UWA channels causing a degradation in the performance of digital communications: transmissions at a high data rate cause many adjacent symbols to interfere at the receiver, requiring sophisticated signal processing to deal with the generated ISI (Inter-Symbol Interference) [5].

Perhaps the most distinguishing property of acoustic channels is the fact that path loss depends on the signal frequency. This dependence is a consequence of absorption, i.e., transfer of acoustic energy into heat. In addition to the absorption loss, signal experiences a spreading loss which increases with distance. The ambient noise, which is always present, may be modeled as Gaussian, but it is not white. Its power spectral density decays at approximately 18 dB/decade. The acoustic bandwidth depends on the transmission distance. The bandwidth is severely limited at longer distances: at 100 km, only about a kHz is available. At shorter distances, the bandwidth increases, but it will ultimately be limited by that of the transducer. The fact that bandwidth is limited implies the need for bandwidth efficient modulation methods if more than a bps/Hz is to be achieved over these channels. Another important observation to be made is that the acoustic bandwidth is centered at low frequencies [5].

The fact that the acoustic bandwidth depends on the distance has important implications on the design of underwater networks. Specifically, it makes a strong case for multi-hopping, since dividing the total distance between a source and destination into multiple hops enables transmission at a higher bit rate over each (shorter) hop. The same fact helps to offset the delay penalty involved in relaying. Since multi-hopping also ensures lower total power consumption, its benefits are doubled from the viewpoint of energy-per-bit consumption on an acoustic channel [5].

By dividing the available bandwidth into a number of narrower bands, Orthogonal Frequency Division Multiplexing (OFDM) [7] systems can perform equalization in frequency domain and eliminate the need for complex time-domain equalizers. OFDM modulation and de-modulation can easily be accomplished using Fast Fourier Transforms (FFT) [7]. When the delay spread is long, the prefix length can significantly affect the bandwidth efficiency. Maximum likelihood sequence detection (MLSD) on individual sub-carriers using a low complexity PSP can combat ISI when the symbol period is smaller than the delay spread. Other channel shortening techniques such as sPRE may also be used in future OFDM systems to reduce the prefix length and improve bandwidth efficiency. Careful consideration of the physical layer parameters can help to design data packets so as to take maximal advantage of limited resources.

Low carrier frequencies typically 15 kHz is ideal for underwater acoustic data communication. Acoustic waves experience $1/R^2$ attenuation due to spherical spreading and also absorptive losses [5] are significant in underwater scenario. Hence, the practical range selected for our carrier frequency is 11-19 khz.

A. Ad hoc Networks

A wireless ad hoc network is a system of self-directed nodes which form a decentralized communications network. Wireless communication allows for a dynamic network topology where new nodes can be rapidly deployed and likewise rapidly removed. The nodes act as both host and router, performing tasks and forwarding information to each other. The mobile nodes can form dynamic networks where they are linked with their nearest neighboring node and when they move too far from their neighboring nodes might lose connection but come into contact with other nodes to begin interacting and changing the network topology. Efficient routing protocols is needed to communicate new data over multi-hop paths consisting of possibly several links to cope with noise and interference as well as sharing limited bandwidth. A class of Ad hoc networks, UANET are used in underwater explorations.

1) UWSN and UANETs: UANET and UWSNs are essential to explore large uninhibited oceans. In the characteristics of these new networks, the propagation delay, floating node mobility, and limited acoustic link capacity are hugely different from ground based mobile ad-hoc networks (MANET) and Wireless Sensor Networks (WSN). UANET and UWSN rely on low-frequency acoustic communications because RF radio does not propagate well due to underwater energy absorption. Unlike wireless links amongst land-based ad hoc nodes, each underwater acoustic link features large-latency and low bandwidth. Most ground sensor nodes in a WSN are typically stationary and large portion of UWSN sensor nodes, except some fixed nodes mounted on the sea floor are with low or medium mobility (3-5 knots) due to environmental water current. The large-scale aquatic applications demand to build UANET and UWSN to explore the large uninhabited oceans. The difference between UANET and UWSN is due to controlled mobility and associated implementation cost. In a UANET, mobile nodes can be implemented by SAUV and Autonomous Underwater Vehicles (AUV) or Remotely Operated Vehicles (ROV), which are high cost robots that can move under the water by following pre-programmed or autonomous motion patterns. On the other hand, UWSN only incurs a fraction of implementation cost of UANET at the same network scale. All sensor nodes in a UWSN are of low-cost [8].

 ISBN: 978-1-61208-296-7

The advantages of the new UANET and UWSN paradigm are:

- Localized and coordinated sensing and attacking is far more precise than the existing remote telemetry technology.
- Scalability of UWSN ensures that a large area can be covered for time-critical applications.
- Casualty ratio is expected to be zero if unmanned UANET and UWSN platforms are used.
- Implementing reusable underwater nodes reduces the deployment and maintenance cost. Each underwater sensor unit can be bundled with an electronically controlled air bladder device. Once the network mission is accomplished, the command center issues commands to trigger all air-bladder devices and all sensor units float to surface to be recollected for next mission [1].

VI. GAF Protocol

When ad hoc networks are deployed using battery powered sensor nodes, the effect of limited battery power on the lifetime and performance becomes critical. In underwater applications, it is vital to let every underwater node know its current position and the synchronized time with respect to other coordinating nodes. GAF protocol uses Global Positioning System (GPS) to get the node location.As Global Positioning System (GPS) is unavailable under the water surface as the high-frequency radio waves used by Global Positioning System (GPS) is quickly absorbed by water, hence cannot propagate deeply under the water surface. Therefore, underwater networks rely on Doppler Instrumentation or distributed GPS-free localization and time synchronization schemes to let the sensor nodes know their positions and the network clock value. In other words, before the network can use geo-routing schemes, it needs a multi-hop packet delivery service, which must be GPS-free. Geographic adaptive fidelity protocol is an energy effective position based routing protocol. Position based protocols are also referred to as geographic routing protocols as the sensor nodes are addressed by means of their locations instead of the information that they carry. Location information is needed in order to calculate the distance between two particular nodes so that energy consumption can be estimated. In GAF protocol, each node uses location information to associate itself with a virtual grid so that the entire area is divided into several square grids, and the node with the highest residual energy within each grid becomes the master of the grid. Only a single node from a cell of a given virtual grid is chosen to be active at any given time. The nodes will select one sensor node to stay awake for a certain period of time which is responsible for monitoring and reporting data to the sink on behalf of the other nodes in the zone is known as the master node. Other nodes in the same grid can be regarded as redundant with respect to forwarding packets, and thus they can be safely put to sleep without sacrificing the routing fidelity [3].

A. GAF Architectures

- Virtual Square Grids: In GAF, the entire network area is divided into virtual square grids. All the nodes in the network divide themselves in virtual square grids and all those nodes which are under a same grid known as equivalent nodes with respect to forwarding packets. The nodes under a same grid coordinate among themselves to decide the sleep time interval and sequence of sleep. Load balancing is performed and a single node will not get drained with a rigorous work.

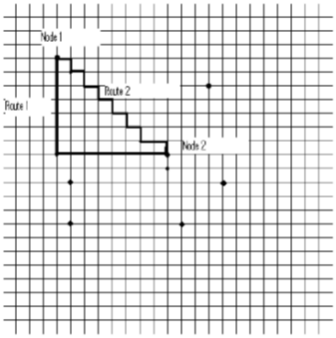


Fig. 4. GAF square grid structure

GAF algorithm can communicate directly to its adjacent horizontal and vertical grid cells. But the diagonal cell cannot be covered directly by virtual grid method due to range limitations. For the diagonal cell, packets should be transferred through vertical and horizontal cell which cause the longer path else if a node tries to transmit directly then the packet drop rate increases due to low radio range. In the figure illustrating the virtual square grid, the Node 1 is sending packets to Node 2. It can take any route vertically and horizontally (see Fig. 4).

- Virtual Hexagon Grid: The hexagon GAF grid architecture uses the hexagonal grid structure. In this, the square grid in GAF is replaced with a hexagon mesh. Cell O now has six neighbors covering destinations from all directions. The hexagon architecture is named as GAF-HEX (HGAF). A Hexagon cell in GAF-HEX is defined as, for two adjacent cell O and B, all nodes in cell O can communicate with all nodes in cell B and vice versa. For a cell O, all of its six adjacent cells are at the next hop, they have the same maximum distance to cell O (see Fig. 5).

  This is the scoring feature of hexagon mesh. In the square grid architecture, there are eight neighboring cells (four diagonal, two vertical and two horizontal cells) but only four (vertical and horizontal two each) are at the next hop distance while the hexagon cell covers all six possible next hop cells with a single maximum distance due to its symmetry property. Therefore all of the next hop cells

 ISBN: 978-1-61208-296-7

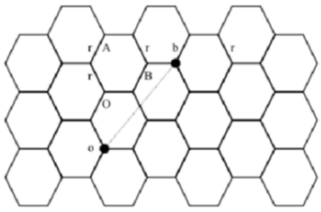


Fig. 5. GAF hexagon structure

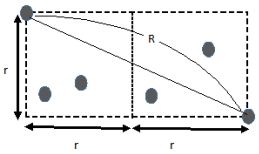


Fig. 6. GAF virtual grid

for cell O are equally reachable.

GAF, the basis of HGAF, is an adaptive fidelity algorithm in which nodes working as cluster heads are selected in a distributed manner. GAF assumes that a large number of sensor nodes are placed in the observation area. The fewest nodes in the observation area are selected to transmit messages, while the other nodes sleep. This way, GAF reduces the number of nodes needed to form a network and saves node battery power. GAF divides the observation area into square/hex areas and groups nodes according to their position. In each group, one active node is selected to work as the cluster head for routing packets between groups. Other nodes sleep to save their battery power. The selection of the active node in each group is done in a distributed manner by referring to the remaining battery power of each node [9].

Assume that the observation area is divided into squares with r units on a side. We call each square a cell. Each node decides which cell it belongs to according to its position. One active node is selected in each cell. Thus, if we can enlarge the cell size, we can reduce the number of active nodes in the network and save even more battery power. However, in enlarging the cell size, the communication between active nodes in two adjacent cells must be guaranteed because active nodes work as cluster heads. Therefore, the distance between the two farthest nodes in any adjacent cell has to be smaller than the radio communication range. The maximum cell area SGAF is constrained by the maximum radio communication range of sensor nodes R. The length of each cell r has to satisfy the following condition:

$$r^2 + (2r)^2 \leq R^2$$

Thus, $r \leq \frac{R}{\sqrt{5}}$

The maximum cell area is thus

$$S_{GAF} = (max r^2) = \frac{R^2}{5}$$

In GAF, the nodes are in one of three states: sleeping, discovery, or active. Nodes start out in the discovery state. One node in each group is selected as an active node, by referring to the residual battery power of nodes in the group. The active node in each group is changed dynamically as time passes. When a node is in the discovery state, it turns on its radio and broadcasts a discovery message to find other nodes within the same group. Each discovery message includes a node ID, group ID, estimated node active time, and the node state. Half of the estimated node active time is denoted by Ta. When a node enters the discovery state, it sets a timer which expires Td seconds later. When the timer goes off, the node broadcasts a discovery message and becomes active. Then, it sets a timeout value Ta that specifies the length of time in which it can stay active. After Ta, the node returns to the discovery state. While in the active state, the node periodically broadcasts its discovery message at intervals of Td. A node in discovery or active states can change to the sleeping state when it receives a discovery message.

- When a discovery node receives a discovery message, it goes into the sleeping state.
- When an active node receives a discovery message, the node compares its own expected lifetime with the one in the discovery message. If the former is longer than the latter, the node goes into the sleeping state.

A sleeping node wakes up after Ts seconds and goes back into the discovery state. Ts is calculated from the estimated node operation time of the current active node, which is written in the received discovery messages [9].

VII. System Architecture

The general architecture of underwater sensor network is reviewed before describing the specific applications. The rough capabilities of a sensor node are estimated on its interaction with the environment, other underwater nodes and applications. At the lowest layer is the large number of sensor nodes to be deployed on the sea floor which has computing power, and storage capacity. They collect information through their sensors and communicate with other nodes through short-range acoustic communication. In large networks, there exists a type of nodes, called

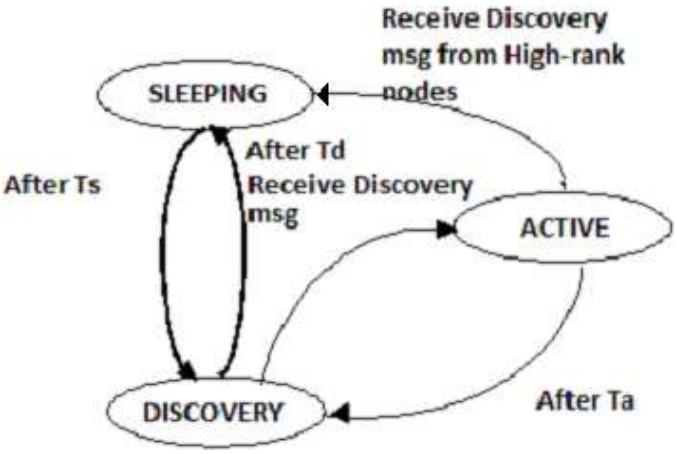


Fig. 7. GAF transition states

supernodes, having access to higher speed networks and can relay data to the base station very effectively with rich network connectivity creating multiple data collection points. Battery power and the ability to carefully monitor energy consumption are essential for the sensor node. All components of the system operate at as low a duty cycle as possible which is enabled by examining each layer of system software to minimize energy consumption and in addition nodes entirely shut off for very long periods of time, up to hours or days during not in use. In a harsh, underwater environment, some nodes will be lost over long deployments due to fishing trawlers and underwater life affecting cables or node which needs redundancy in communication and sensing as loss of a node will not have wider effects. In addition, multiple failures can be recovered, either with mobile nodes or with human deployment of replacements [11].

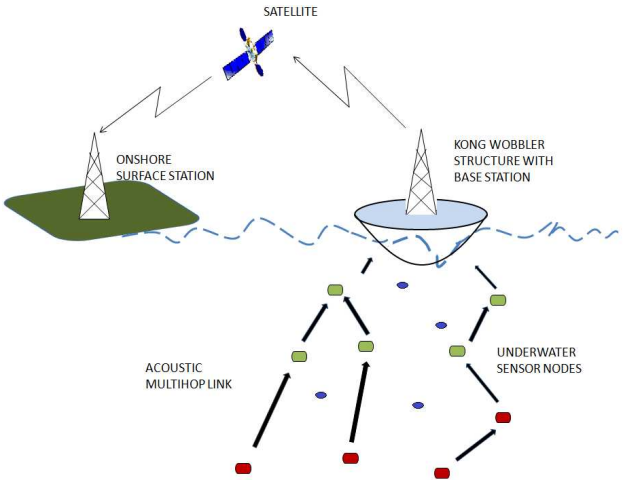


Fig. 8. Logical diagram of System Architecture.

VIII. Implementation

Initially, the UWSNs and UANETs are deployed in the ocean. Sensor nodes with limited battery power are deployed to record the environmental changes underwater. The recorded data is transferred to the surface of the earth through the nearest access point, the base station fixed to the specially designed Kong Wobbler floating on the surface of the water by the multi-hop network of sensor nodes which results in energy savings and increased network capacity. The base stations monitor the entire sensor nodes within the network and receive as well as stores the recorded sensor data within its area of range. These data are further relayed to an onshore surface station via satellite transceiver. Acoustic communication proves to be efficient in data transfer underwater from the sensor nodes to the base station. During the process, the base station monitors the battery power level of all the nodes. Upon receiving the request from the nodes to recharge its batteries, the base station then guides the submarines towards the requested nodes which is continued with all the sensor modules deployed underwater. The sub-marines would in turn charge themselves at the base stations, which are installed with the solar panels. The power required by the base station for the reception of the data from the sensor modules and transmission is supplied by the solar panels. Once the surface station has finished collecting data from each sensor module, the processing and analysis of data is performed to get the real-time study of the underwater scenario. As the nodes have limited battery power, it is essential to implement an energy efficient routing protocol that conserves power during transmission and reception of data. Power failure of a node not only affects the node itself but also its ability to forward packets on behalf of other nodes and thus the overall network lifetime. GAF is an energy efficient routing protocol as the transmitting power is altered according to the distance of the neighboring nodes. The flow chart explains the process undertaken in the investigation of underwater environment.

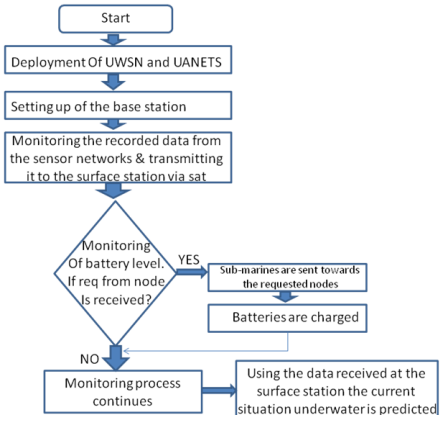


Fig. 9. Flow Chart of the system

The investigation of the underwater resources is thus col-

lected at real-time. Compilation of all the recorded sensor data is done in the surface station which acts as the command center and thus predicting the current scenario of the environment underwater.

IX. Simulation and Results

Continuous network development and higher functionality requirements have created the need for tools that could monitor network transmissions and analyze them. Network Simulator (NS) for communication networks works under UNIX and Windows system platforms and is mainly used for network research. The simulation is performed using NS2 (version 2.34) running on LINUX platform (Ubuntu 11.04). The graphical representation of this simulation is shown with Network Animator(NAM-1.14). 20 nodes are considered for each Kong Wobbler base station on the surface of ocean bed. The Kong Wobbler carrying base station act as the center that monitors the entire sensor node within the network and receives as well as stores the recorded sensor data. UANETs along with Kong Wobbler structure is hand dropped by the side of the ship. Kong Wobbler structure starts to float on the surface of the ocean and the UANETs start moving randomly. The black fixed nodes are anchored to the ocean bed at specific co-ordinates (see Fig. 6). Some of the nodes in UANET are mobile which are implemented by SAUV and AUV or ROV, which are high cost robots that move under the water by following pre-programmed or autonomous motion patterns. And remaining nodes in UWSN are stationary, mounted on the sea floor are with low or medium mobility (3-5 knots) due to environmental water current. The trace file and NAM file results provided by the NS2 gives enormous amount of information. It specifies position of the node, number of nodes within the network of access point and also visualizes in detail about the packet transmission among the nodes and the Kong Wobbler carrying base station is simulated. The red nodes represent nodes in UANET on the ocean bed which are moving around collecting data. The black nodes are fixed nodes anchored to oceanic bed. The blue node indicate the Kong Wobbler structure with base station which covers the area under investigation.

Nodes discovering their themselves (Discovery State), (see Fig. 10)

UANETs continue their random motion gathering data. Both stationary UWSNs and mobile UANETs collect information and relay it to base station which is carried by Kong Wobbler structure (see Fig. 11). The acknowledgements are sent after receiving the information.

NS simulation can produce a visualization file (NAM) and an ascii file trace corresponding to the events generated in the network. When tracing into an output ascii file, the trace is organised into 12 fields as- Event/Time/From Node/To Node/Pkt Type/Pkt Size/Flags/Fid/Source addr/Destination addr/Sequence no/Pkt ID. The Traces are studied and information is extracted from them to make Performance analysis. Trace graph architecture enables implementing new system

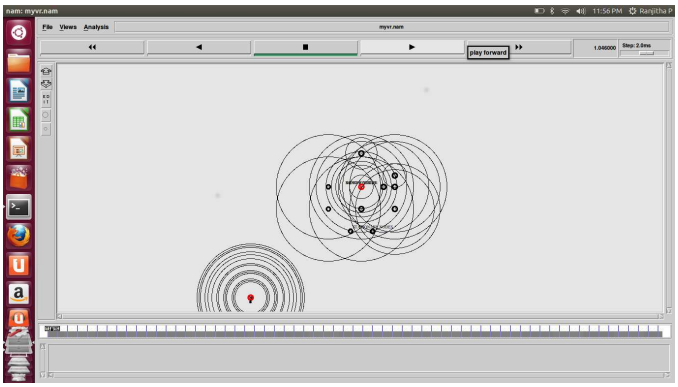


Fig. 10. Simulation in NS2

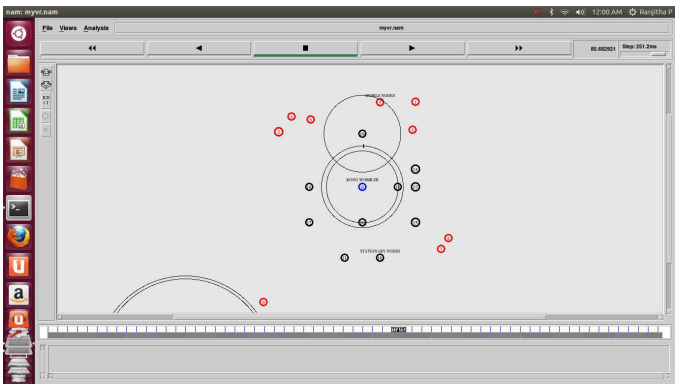


Fig. 11. Simulation in NS2

functions very easy. For example adding a new graph can take only 10 minutes. The system could be expanded to read other trace file formats like real network traces, e.g., a trace format converter could be created for conversion to Trace graph format. Nodes movements with packets flows 2D/3D visualization could be added. More parameters from trace files could be used for new graphs implementation; see snapshot of our Trace file in Fig. 12.

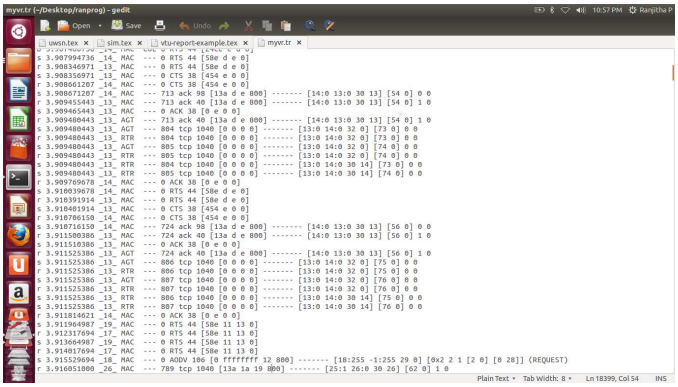


Fig. 12. Trace file

The investigation is carried out in two phases considering: Packet Delivery Fraction (PDF): It is the ratio of the data packets delivered to the destination to the total number of

 ISBN: 978-1-61208-296-7

packets generated by the Constant Bit Rate (CBR) source

1) First phase is with variation of the speed of the nodes with PDF for AODV-GAF and DSDV-GAF is as shown in Fig. 13.

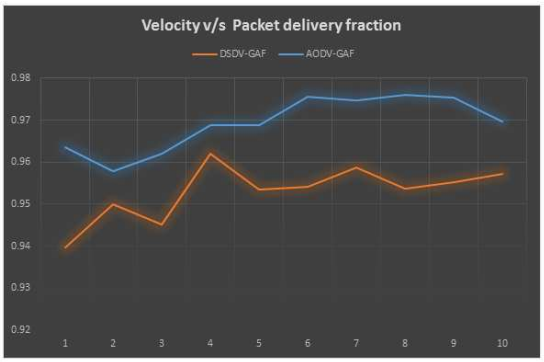


Fig. 13. Velocity v/s PDF

2) Second phase is the variation of simulation time with PDF AODV-GAF and DSDV-GAF is as shown in Fig. 14.

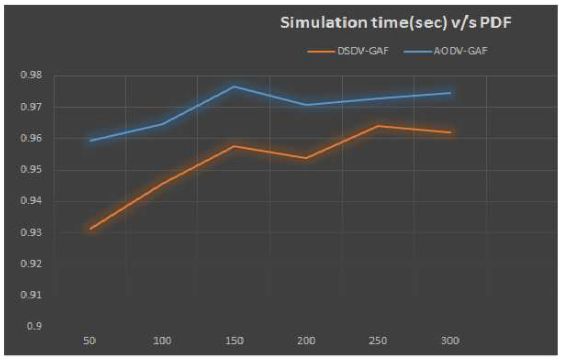


Fig. 14. Simulation Time v/s PDF

Inferring from the graphs, GAF implemented on AODV protocol is more suitable for underwater environment than DSDV-GAF.

X. Conclusion

This paper has summarized our ongoing research in underwater sensor networks, including applications and research challenges. It is explained that traditional approach to deploy underwater sensors that record data during the monitoring mission, then recovering the instruments is not a feasible and the need of large-scale long-term and distributed information collection networks for periodic oceanic monitoring is essential. GAF (Graphical Adaptive Fidelity) protocol was adopted as it proves to be an energy efficient routing protocol. It is also explained that acoustic communication is the most versatile technique in underwater wireless communication. The applications of UANET using SAUV and UWSN and their reliability in implementing a localized, precise, and large-scale networking efficiently than any existing small-scale Underwater Acoustic Network (UAN) is described. The main objective of the paper was to develop advanced communication techniques for efficient real-time investigation of large uninhibited oceans. Development of underwater communication and networking for enhanced oceanic monitoring is also essential for pollution monitoring, tactical surveillance, exploration of natural undersea resources, predicting wave tides and various applications.

References

[1] Ian F. Akyildiz, Dario Pompili, and Tommaso Melodia, "Challenges for efficient communication in underwater acoustic Sensor Networks", ACM SIGBED Review - Special issue on embedded sensor networks and wireless computing, Homepage archive, vol. 1, issue 2, July 2004, pp. 3-8. ACM New York, NY, USA , http://10.1145/1121776.1121779.

[2] Eitan Altnan and Tania Jinenez, "NS simulator for beginners", Lecture notes 2003-2004.

[3] Swarnalatha Srinivas, Ranjitha P, R Ramya, and Narendra Kumar G, "Energy efficient investigation of oceanic environment using large-scale UWSN and UANETs", IJCSI International Journal of Computer Science Issues, vol. 10, issue 1, no. 1, January 2013, pp. 566-573

[4] Robert H. Stewart, "Introduction to physical oceanography".

[5] Ian F. Akyildiz, Dario Pompili, and Tommaso Melodia, "Underwater acoustic sensor networks: research challenges", Telecommunications (ICT) 2010 IEEE 17th International Conference, pp. 257-279, http://dx.doi.org/10.1016/j.adhoc.2005.01.004.

[6] James Jalbert, John Baker, John Duchesney, Paul Pietryka, William Dalton Acoustikos Div. Falmouth Scientific, Inc., et al, "Solar-Powered Autonomous Underwater Vehicle Development", supported by the Office of Naval Research under Contract # N00014-03-C-0109.

[7] Milica Stojanovic, "Underwater acoustic communications: Design considerations on the physical layer", Wireless on Demand Network Systems and Services (WONS), 2008, Garmisch-Partenkirchen, pp. 1-10, http://10.1109/WONS.2008.4459349.

[8] Jiejun Kong, Jun-hong Cui, Dapeng Wu, and Mario Gerla, "Building underwater Ad-hoc networks and sensor networks for large scale real-time aquatic applications", IEEE MILCOM, 2005, Atlantic City, NJ, vol. 3, pp. 1535-1541, http://10.1109/MILCOM.2005.1605894

[9] Tokuya Inagaki and Susumu Ishihara, "HGAF: A power saving scheme for Wireless Sensor Networks", IPSJ Journal, vol. 50, no. 10, Oct 2009, pp. 25202531.

[10] Jun-Hong Cui, Jiejun Kong, Mario Gerla, and Shengli Zhou, "Challenges: Building scalable and distributed Underwater Wireless Sensor Networks (UWSNs) for aquatic applications", UCONN CSE Technical Report: UbiNet-TR05-02, Last Update: September 2005, pp. 1-17.

[11] John Heidemann, Yuan Li, Affan Syed, Jack Wills, and Wei Ye, "Underwater Sensor Networking: Research Challenges and Potential Applications", The IEEE Wireless Communications and Networking Conference, Las Vegas, Nevad, USA, April 2006.

[12] Sudhakar Pillai M, Pranav P D, and Narendra Kumar G, "MANET based Dynamic Power Conscious emergency communication module", 01/2011; In proceeding of: LISS 2011 - Proceedings of the 1st International Conference on Logistics, Informatics and Service Science, vol. 3, Beijing, China, 8-11 June, 2011.

[13] Giuseppe Anastasi, Marco Conti, Mario Di Francesco, and Andrea Passarella, "Energy conservation in Wireless Sensor Networks: a Survey".

[14] Robert Been, David T. Hughes, and Arjan Vermeij, "Heterogeneous underwater networks for ASW: technology and techniques", UDT Europe, Undersea Defence Technology Europe, Glasgow, United Kingdom, 10-12 June, 2008, pp. 1-17.

[15] Lanbo Liu, Shengli Zhou, and Jun-Hong Cui, "Prospects and problems of wireless communication for Underwater Sensor Networks", Journal: Wireless Communications & Mobile Computing - Underwater Sensor Networks: Architectures and Protocols, pp. 1-25.

 ISBN: 978-1-61208-296-7

[16] Raja Jurdak, Cristina Videira Lopes, and Pierre Baldi, ”Battery lifetime estimation and optimization for Underwater Sensor Networks”, IEEE Sensor Network Operations, IEEE Press, Winter 2004, pp. 1-25

[17] Sudhakar Pillai M, Pranav P Deshpande, Chetan B M, Smitha Shekar B, and Narendra Kumar G, ”Efficient performance Of MANETs in coal mines”, Unpublished.

[18] Swarnalatha Srinivas, Ranjitha P, R Ramya, and Narendra Kumar G, ”Investigation of oceanic environment using large-scale UWSN and UANETs”, The 8th International conference on Wireless Communications, Networking and mobile computing, Shanghai, China, sep 21-23, 2012, pp. 1-5, http://dx.doi.org/10.1109/WiCOM.2012.6478552.

 ISBN: 978-1-61208-296-7

Ocean Space Surveillance - Network Beployment Based on Hydrodynamic Modeling

Tor Arne Reinen,
Knut Grythe
SINTEF
Trondheim, Norway
Tor.A.Reinen@sintef.no
Knut Grythe@sintef.no

Dag Slagstad,
Morten Omholt Alver
SINTEF Fisheries and Aquaculture
Trondheim, Norway
Dag.Slagstad@sintef.no
Morten.Alver@sintef.no

***Abstract*—The use of hydrodynamic modeling to optimize the placement of sensor nodes with respect to reduction of uncertainty in sea current modeling is demonstrated. This is done under the additional constraint that the nodes should constitute a connected underwater acoustic communication network. For the connectivity analysis a sound propagation model is employed, based on detailed sound speed data produced by the hydrodynamic model. It is found that such a detailed sound speed field is very attractive for obtaining sufficient precision in acoustic modeling. This is particularly important in regions of significant variability in bathymetry and sea currents. In such regions it would be very demanding to obtain sufficiently detailed data by physical measurement, and the use of a hydrodynamic model is virtually the only alternative. In general the model based approach is important for up-scaling the geographical extent of a network, and has a large impact on the economy of its deployment and operation.**

Keywords— hydrodynamic modeling; underwater acoustic network; network deployment; ray tracing;

I. INTRODUCTION

Under the constraint of limited number of nodes in a sensor network it is of high importance to optimize the placement of the nodes. Underwater acoustic sensor networks typically need to obey such limitations, due to high costs of sensors, communication modems and of deployment and maintenance operations. To optimize with respect to the measurement task, knowledge about the process is needed. To optimize with respect to communication, knowledge about sound propagation conditions is needed. The present paper aims to demonstrate the use of hydrodynamic modeling to guide deployment based on both of these criteria together. This is carried out by a simple case study. Though the availability of hydrodynamic model tools is quite widespread, this opportunity seems to have been exploited quite scarcely in the literature, in particular for the joint measurement and communication design.

Our work contributes to the development of an ocean space surveillance concept [1], i.e., to combine underwater sensor networks and ocean modeling in order to obtain improved surveillance capabilities with respect to both real-time situation awareness and model based predictions and assessments. The system architecture is composed of an underwater sensor network, communication links to computers running ocean models, and the ocean models themselves. The latter assimilate measured data from the sensor network in a manner similar to what is done in meteorology.

The measurements that we want to optimize are those that will be used for data assimilation: Time series of current velocity vertical profiles, obtained by upwards-looking profiling sensors close to the sea floor, at geographically fixed locations. In regions of variable bathymetry and external currents, especially coastal region, the process (the current) is very far from uniform in space and time.

In our realization of the concept, and in many related scenarios, cost issues lead to strict limitations in the number of underwater nodes. An optimal deployment is therefore essential to obtaining a sufficient system performance.

There is a considerable body of publications on sensor array design for meteorology and metocean purposes. Our work is based on [2], with extensions according to [3]. Good further references can be found in [2].

Sound propagation is mainly governed by the spatial variability of the speed of sound, together with boundary conditions at the sea surface and bottom. Sound speed variability causes sound rays to curve in space, which may enhance or preclude propagation. In our case, sound propagation translates to connectivity between sensor nodes. Many references consider modelling of propagation based on the assumption that the sound speed field is known [4][5]. This is, however, often only approximately true, and the resulting lack of precision can produce unfortunate deployments if the number of nodes must be limited, with lack of connectivity as a result. Solving the sound speed data issue by using hydrodynamic models seems scarcely studied, in particular in connection with network technology. A few references investigate general sound speed variability prediction from hydrodynamic models [6][7], in [7] including acoustic ray tracing as a means to evaluate the modeling output. Reference [8] combines local hydrodynamic modeling and a parabolic equation model for sound propagation. Reference [9] is an example of the inverse problem: Assimilating tomographic sound propagation measurements in hydrodynamic models. Hardly any references seem to use hydrodynamic models in the context of network connectivity. This is not fundamentally different from the above examples, but the focus largely determines how the hydrodynamics results are exploited. In [10] we consider this topic, but do not include the issue of where to place nodes for optimal benefit of the sensor measurements.

Joint deployment design of sea current sensors and communication network is investigated in [11], but only local

 ISBN: 978-1-61208-296-7

sea current dynamics and a uniform bathymetry are considered. In the present paper a full scale hydrodynamic model is introduced, in combination with detailed bathymetry information. The subsequent sound propagation analysis is considerably influenced by these inputs.

The paper is organized as follows: Section II describes our deployment-design strategy, including hydrodynamic modelling, optimization of node localization and sound propagation modelling. Section III presents our case-study and section IV the corresponding simulation results. Finally Section V gives a summary and conclusions from our work.

II. DEPLOYMENT-DESIGN STRATEGY

Our strategy is entirely model based: First the hydrodynamic model is used to (i) identify sensor node locations and (ii) provide detailed spatial and temporal sound speed fields. The latter are then input to an acoustic propagation model that evaluates connectivity of the underwater acoustic communication network. If necessary, node locations are slightly modified to ensure connectivity. If larger position-modifications are necessary, relay nodes are added (it is then an issue of incremental benefit and cost whether or not to include sensors also at such locations).

A. Hydrodynamic modeling

Our hydrodynamic model (SINMOD) is based on the primitive Navier-Stokes equations and is established on a z-grid [12]. Each model level has a fixed thickness except for the surface levels and the one close to the bottom. Turbulent vertical mixing is calculated as a function of the Richardsons number and horizontal mixing is calculated according to [13]. The model uses a nesting technique to allow high resolutions in targeted areas. Usually a basin scale model setup having coarse horizontal grid resolution (~20 km) in nested in several steps down to 30-50 m. Usually 8 tidal components are forced on the open boundaries of the basin scale model. The models are then forced by atmospheric fields (wind, pressure and heat flux) and river run-off.

High resolution versions (800 and 160 m) of the model have been applied for various parts of the Norwegian coast. Skill assessment of the model against a data set consisting of 9 Acoustic Doppler Current Profiler moorings placed on the Vesterålen-Lofoten shelf showed that the model was doing well [14].

B. Optimization of node localization

Optimization of the localization of the measuring nodes is carried out so as to minimize uncertainty of the states (current field) of the hydrodynamic model, when including measurement input. This criterion is integrated over a chosen region of interest. Localization of the impact of each sensor is enforced, i.e., influence on geographically distant states is blocked. The procedure starts with a greedy search based on [2], i.e., the nodes are optimized in sequence, at each step considering the others to be fixed. The procedure is iterated, but inevitably produces a suboptimal result. The final output is improved by a tabu search according to [3]. In essence, this strategy systematically searches in the neighbourhood of a sub-optimal solution to find the best measurement locations.

C. Sound propagation

Sound propagation is modeled by ray tracing, This is the preferred method for frequencies of interest for acoustical communication, above some 10 kHz. Rays have been calculated by direct integration (see [4] ch.3.5). Amplitudes along rays have similarly been directly integrated by the simplified dynamic ray tracing equations of [15]. This is in effect a classical ray tracing system, including (non-physical) infinite amplitudes at caustics. The caustics artifacts are not removed from the results, but caustics are detected in each ray and an ad hoc number of range steps (10) around them are removed from the results. Boundary conditions have been implemented in accordance with [16]. Hence, scattering losses at the rough boundaries are generally lower than the classical coherent reflection coefficient. This, however, means that the reflected arrivals at the receiver are a time spread sum of incoherent contributions.

Rays are terminated when a maximum selected transmission loss (TL) is reached. This termination is the only amplitude information in the present paper: TL is less than the chosen maximum everywhere on the plotted rays. For the present simulations, the selected TL maximum is chosen from a simple transmission budget assessment (corresponding to the sonar equations of [17]): With a source level SL and a receiver noise level NL, the signal to noise ratio at the receiver is given by SNR = SL – TL – NL. The SNR is then required to be at least equal to a detection threshold, DT. The transducers used are assumed to be omni-directional due to the orientation robustness usually required in sensor networks. At 1 m reference distance this gives $SL = 10\log_{10}(\rho c P/4\pi)$, where P is acoustically radiated power, c is sound velocity and ρ is the medium density. Expressed in dB re. 1 μPa rms sound pressure, this amounts to $SL = 170.8 + 10\log_{10}(P)$. The noise level is $NL = 10\log_{10}(N_0 B)$, where N_0 is the spectral density of the ambient noise and B is the system's noise bandwidth. Here omni-directional transducers are again assumed. For the present simulations the system is assumed to operate at 12.5 kHz, with B = 300 Hz and N_0 = 48 dB re. 1 μPa/√Hz. The latter corresponds to open sea wind generated noise at Sea State 5 [17]. Furthermore, P = 20 W and DT = 20 dB are assumed. These parameters, then, give a maximum TL of 90 dB (and this value is reached at range 7.7 km for the uniform free space case). The same TL requirement would result for any combination of P, B and N_0, as long as $P/(BN_0)$ is constant.

III. THE CASE STUDIED

The case selected for our study is the area around the Tristeinen islets at the coast of Trøndelag, Mid Norway, see Figure 1.

 ISBN: 978-1-61208-296-7

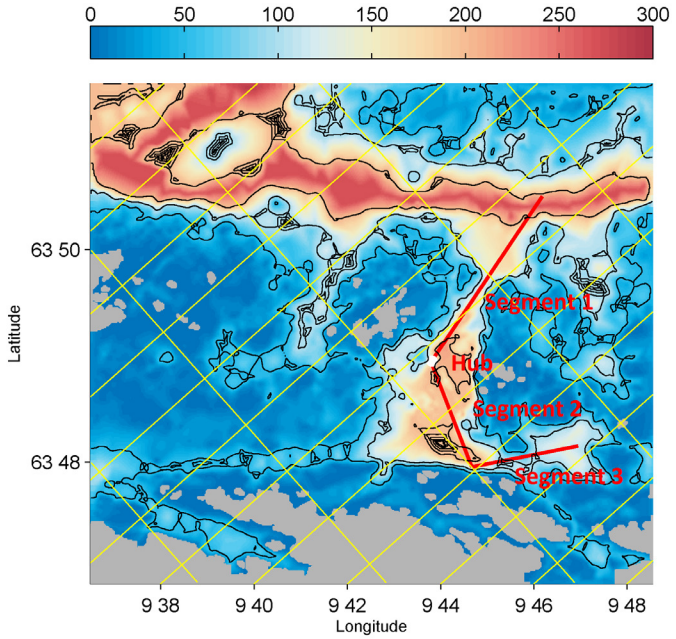


Figure 1 Bathymetry of the deployment area. The color scale shows depth (m). Gray areas are dry land. The yellow coordinates are true north and east. Axis labels refer to these coordinates. The model grid orientation is rotated 48 degrees in relation to true north. Text in red refers to a network deployment case of which Segment 1 is studied in the present paper.

The area is inside the main coastal currents and the currents are relatively weak. This has been confirmed in measurements as well as in the model simulations. The coastal water warms up in the surface during the summer and increases the sound speed compared to deeper waters. In autumn and winter when strong cooling of surface water takes place, the sound speed will decrease towards the surface. Low salinity water near the surface amplifies this cooling as vertical mixing is inhibited.

The location has been used as a test site for aquaculture in exposed areas. The sensor network deployment investigated in the present paper is part of a potential extension of the experimental setup described in [1]: We study in some detail Segment 1 of the two-branch (three segment) configuration in Figure 1. One end of the segment is the hub of the two-branch network. This is identified as $r = 0$ in the following. The other end is 5.2 km away, beyond a shallower saddle point approximately at the midpoint of the segment.

IV. SIMULATION RESULT

A. Sensor node placement optimization.

In the present model setup, a basin scale model for the Northeast Atlantic and the Arctic Sea was established with a horizontal resolution of 20 km [18]. High resolution models were nested in 4 steps down to 32 m.

Optimalization of sensor node placement, as described in section II.B is shown in Figure 2. Three nodes were included in the search. White crosses in the figure indicate the output of the greedy algorithm search, while black circles indicate the final locations output from the tabu search (note that the lower-left circle and the upper-right cross overlap). The entire region in the figure was included in the optimization criterion. The result emphasizes the importance of the saddle point region on Segment 1 of Figure 1. This is a region where the basic hydrodynamic model predicts strong currents. The optimization output, then implies that this current is highly correlated with the lower intensity flow in other parts of the region, so that reduction of uncertainty at the sensor locations contribute to a reduction over a wider region.

B. Network connectivity analysis

Sound speed fields were simulated in 3D as described in section IV.A for the period August 21st-December 23rd 2010. Sound propagation analyses based on this were then carried out to evaluate connectivity between network nodes along Segment 1. It was decided to keep only one of the nodes in Figure 2 and place a second one at the far end of Segment 1, in order to capture currents that are of general interest, but that do not contribute significantly to the flow in the region of interest for optimization (not providing further information once the currents at the saddle point are known).

A pair of nodes is considered to be connected if there is at least one ray between them with a TL that does not exceed our chosen maximum. This has not been tested by an explicit eigenray search, but simply by checking whether or not a node is enclosed in a fan of rays transmitted from the other one. Sound propagation analysis for the entire network in Figure 1 can be found in [10], using a somewhat more conservative connectivity definition.

Sound speed time variability at the hub is shown in Figure 3. A clear transition from summer to winter conditions is seen in the middle of the period. Figure 4 shows vertical sections along Segment 1 at the start, approximate midpoint and endpoint of the time series in Figure 3. Here spatial variability in the sound speed profile (SSP) is clearly demonstrated. Note that this would be very demanding to measure physically.

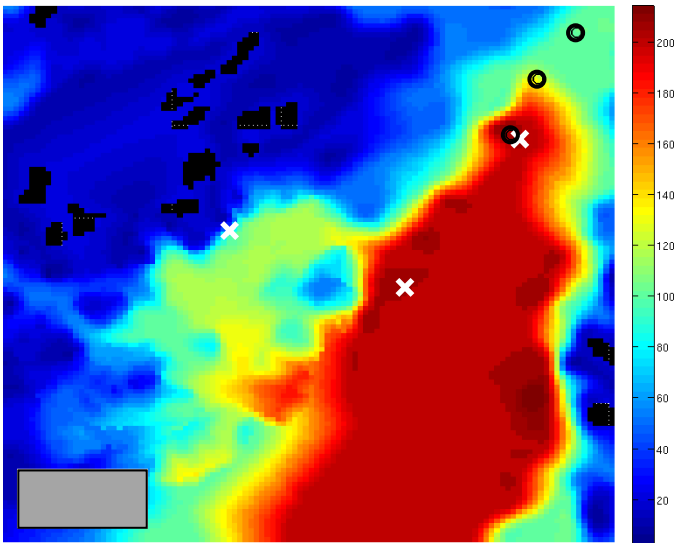


Figure 2 Optimization of sensor node localization: Three nodes included. White crosses are the result of the greedy search. Black circles are final locations after tabu search. Note that the lower-left circle and the upper-right cross overlap. The line from the lower left cross to the upper right circle is approximately the first half of Segment 1 in Figure 1.

 ISBN: 978-1-61208-296-7

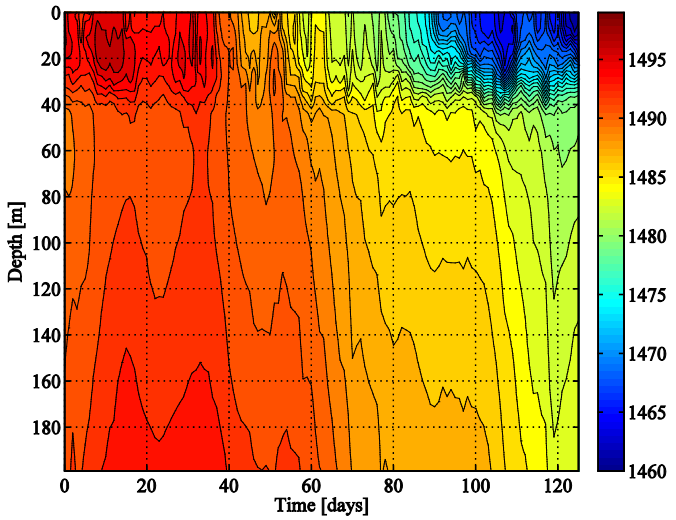


Figure 3 Simulated SSP at network hub. August 21th December 23th 2010. One profile per day

We will now simultaneously demonstrate (i) the use of sound propagation simulations based on modeled sound speed fields – to supplement the above node placement result, and (ii) that very detailed sound speed data are sometimes needed, so that measurement is generally very demanding and the use of modeled input data very attractive.

To this end, vertical sound speed profiles at the hub-location ($r = 0$) in Figure 4 can be claimed to represent either single point local physical measurements or quite good historical data from the same time of the year. Simulations using such profiles can then be compared to simulations using the complete modeled sound speed field, which is much closer to a true sound field. Hence the differences in simulation results illustrate in a conservative manner the advantage gained by using detailed SSP modeling. The comparison is conservative due to the fact that the single range vertical profiles chosen represent quite good data. Much larger differences would be found if generic, season specific data were used.

Now, consider the connectivity between the hub, the optimal saddle point location and the far end of Segment 1 under winter conditions (see the lower part of Figure 4). Figure 5 shows this connection: Ray tracing from the middle node to each end using the vertical sound speed profile at $r = 0$. Both end points are required to be located maximum 5 m above the bottom, while middle node is elevated to the minimum height above the bottom that ensures connectivity to the end points. These requirements are introduced due to mechanical robustness and stability issues. The two magenta lines in the figure show all points located 5 m and 10 m above the sea floor. Connectivity in the directions oposite to those shown is given by reciprocity.

The same ray tracing is repeated in Figure 6, now using the full range dependent sound speed field of Figure 4. We observe that the middle node placement is problematic: There is no connectivity between this node and the far end at 5.2 km.

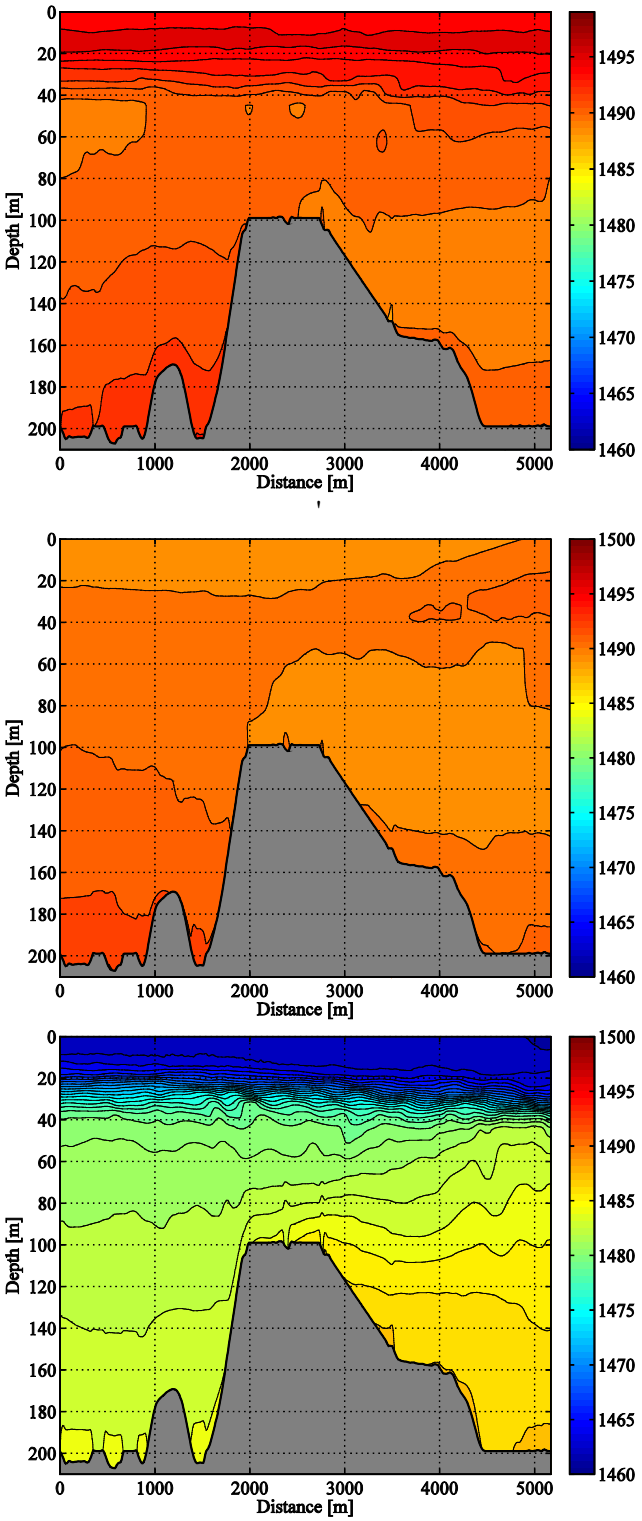


Figure 4 Sound speed range dependent profile. Segment 1. Upper: August 21st. Middle: October 1st.. Lower: December 23rd.

To obtain a connection the middle node is moved around in the vicinity of the position given by the measurement optimization, looking for a position with connectivity. The result is shown in Figure 7, where connectivity in both directions is found.

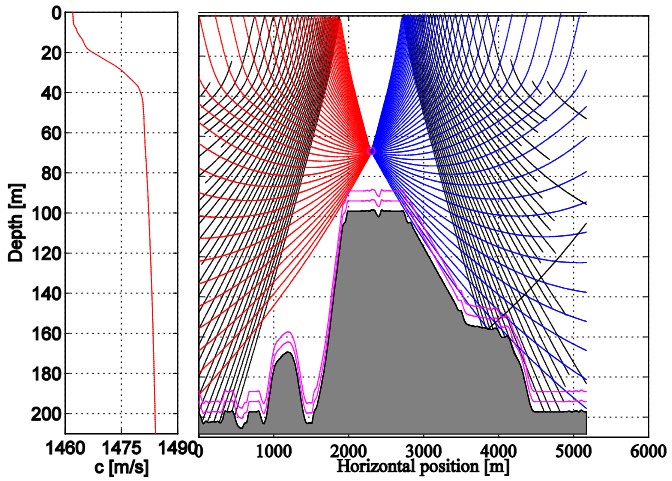


Figure 5 Relay and sensor node 30 m above the bottom. Ray tracing (right) using SSP (left) from hub location, $r = 0$, December 23rd. Colored rays: From relay-towards endpoints, the part before scattering. Black: Ray continuation after scattering in surface or bottom.

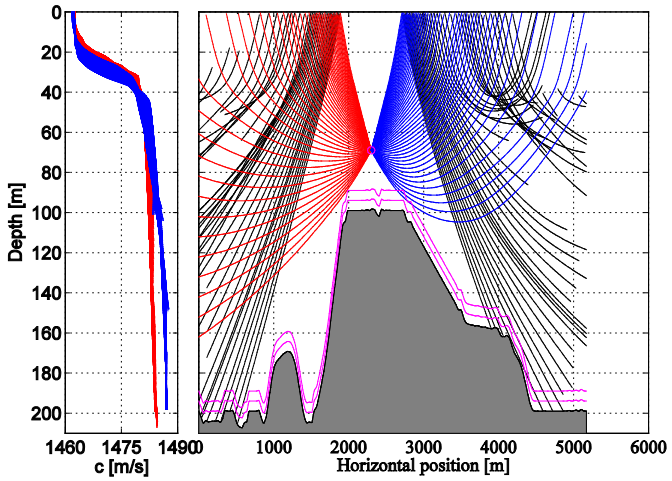


Figure 6 Relay and sensor node 30 m above the bottom. Ray tracing (right) using range dependent SSP (left) – December 23rd. Colors as in Figure 5.

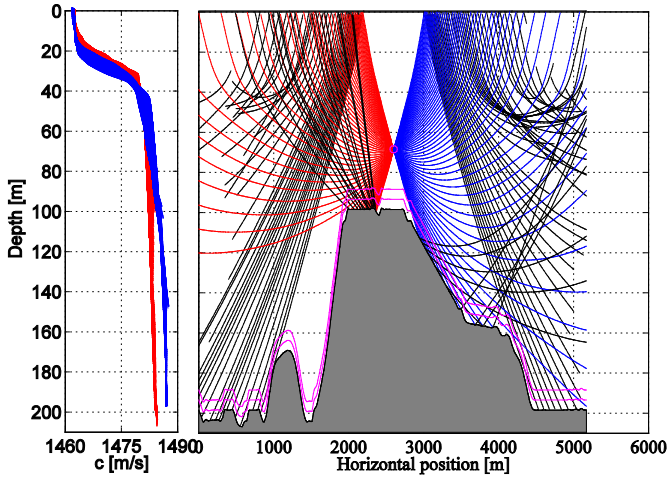


Figure 7 Relay and sensor node moved 300 m right relative to Figure 6. Range dependent SSP from December 23rd. Colors as in Figure 5.

The node in Figure 7 has been moved 300 m away from the start point, a result that is deemed acceptable for the measurement task. The analysis has been repeated for the other two sound speed datasets in Figure 4, with the same positive outcome. Hence, the node location is our final result, and we have demonstrated the two points (i) and (ii) above.

V. SUMMARY AND CONCLUSION

Our case study has demonstrated the use of hydrodynamic modeling to optimize the placement of sensor nodes with respect to uncertainty reduction in sea current modeling, under the additional constraint that the nodes should constitute a connected underwater acoustic communication network. For the connectivity analysis a sound propagation model has been employed, based on detailed sound speed data produced by the hydrodynamic model. Furthermore, it has been demonstrated that such a detailed sound speed field is very attractive for obtaining sufficient precision in acoustic modeling, especially in regions of significant variability in bathymetry and sea currents. It would be very demanding to obtain sufficiently detailed data by physical measurement, and the use of a hydrodynamic model is virtually the only alternative. In general the model based approach is important for up-scaling the geographical extent of a network, and has a large impact on the economy of its deployment and operation.

ACKNOWLEDGMENT

The work has been part Ocean Space Surveillance – OSS, an internally funded strategic research initiative in the SINTEF group, aiming to combine wireless sensor networks and ocean models to obtain high quality real time situation awareness and predictions.

REFERENCES

[1] K. Grythe et. al.,"Ocean space surveillance system – OSS" in Proc. IEEE Oceans'13, Bergen, Norway, June 10-13 2013. ISBN 978-1-4799-0000-8. Paper 130118-063

[2] P. Sakov and P.R. Oke, Objective Array Design: Application to the Tropical Indian Ocean. Journal of atmospheric and oceanic technology, 2008. **25**: pp. 794-807.

[3] F.S. Hillier and G.J. Lieberman, Introduction to operations research. McGraw-Hill, 2005.

[4] F. B. Jensen, W. A. Kuperman, M. B. Porter, and H. Schmidt., Computational Ocean Acoustics. New York: AIP Press, 2000.

[5] M. Stojanovic and J. Preisig, "Underwater acoustic communication channels: Propagation models and statistical characterization," Communications Magazine, IEEE, vol. 47, pp. 84-89, 2009.

[6] M. Sadrinasab and K. Kenarkohi, "Three-Dimensional Numerical Modelling Study of Sound Speed in the Persian Gulf," Asian Journal of Applied Sciences,, vol. 2, 2009, pp. 232-239.

[7] I. Church, J. Hughes Clarke, S. Haig, and R. Toodesh., "Modelling the estuarine circulation of the Port of Saint John: Visualizing complex sound speed distribution.," in Canadian Hydrographic Conference, Niagara Falls, 2012.

[8] S. Finette, T. Evans, and C. Shen, "Sub-Mesoscale Modeling of Environmental Variability in a Shelf-Slope Region and the Effect on Acoustic Fluctuations," in Impact of Littoral Environmental Variability on Acoustic Predictions and Sonar Performance N. Pace and F. Jensen, Eds., ed: Springer Netherlands, 2002, pp 401-408.

 ISBN: 978-1-61208-296-7

[9] J. K. Lewis, J. Rudzinsky, S. Rajan, P. J. Stein, and A. Vandiver, "Model-oriented ocean tomography using higher frequency, bottom-mounted hydrophones," The Journal of the Acoustical Society of America, vol. 117, 2005, pp. 3539-3554.

[10] T.A. Reinen and D. Slagstad, "Connectivity prediction in underwater acoustic networks based on hydrodynamic modeling", in Proc. IEEE Oceans'13, Bergen, Norway, June 10-13 2013. ISBN 978-1-4799-0000-8. Paper 130118-068.

[11] D. Pompili, T. Melodia, and I. F. Akyildiz, "Three-dimensional and two-dimensional deployment analysis for underwater acoustic sensor networks," Ad Hoc Networks, vol. 7, 2009, pp. 778-790.

[12] D. Slagstad and T. McClimans, "Modelling the ecosystem dynamics of the Barents Sea including the marginal ice zone: I. Physical and chemical oceanography. Journal of Marine Systems, 58, 2005, pp. 1-18.

[13] J. Smagorinsky, "General circulation experiments with the primitive equations. I. The basic expriment", Monthly Weather Review, 91, 1963, pp. 99-164.

[14] Annon, "LOfoten and VEsterålen CURents (LOVECUR). Comparison of Hindcast with Measurements". Revison 2.Rep., Forristall Ocean Engineering, Inc, 2011.

[15] M. B. Porter and Y.-C. Liu, "Finite-Element Ray Tracing," in Proc. Theoretical and Computational Acoustics, 1994.

[16] APL, "APL-UW high-frequency ocean environmental acoustic models handbook," Univ Washington, Applied Phys Lab APL-UW TR 9497 APL-UW TR 9407, 1994.

[17] R. J. Urick, Principles of underwater sound for engineers. New York: McGraw-Hill, 1975 (also Peninsula Publishing 1995).

[18] D. Slagstad, I. H. Ellingsen, and P. Wassmann, "Evaluating primary and secondary production in an Arctic Ocean void of summer sea ice: An experimental simulation approach", Progress In Oceanography, 90(1-4), 2011, pp. 117-131.

 ISBN: 978-1-61208-296-7

A Low Cost Turbidity Sensor Development

Sandra Sendra, Lorena Parra, Vicente Ortuño, Jaime Lloret
Universidad Politécnica de Valencia, Spain
sansenco@posgrado.upv.es, loparbo@epsg.upv.es, vortunyo@eln.upv.es, jlloret@dcom.upv.es

***Abstract*— Water turbidity measurement is important for underwater environmental monitoring. Large changes in this parameter can affect fishes catching the food. This avoids having fishes feed properly and most pellets are wasted and deposited in the seabed. In some cases, it is more important to know the variation of the turbidity in real time than a very accurate value. For this reason, we may not need very high precision in our sensor. In this paper, we show the development of a low cost turbidity sensor. Our sensor is able to take measurements in real-time and presents long useful lifetime. Consequently, it will need low maintenance. To test our system, we are going to create a set of samples with different concentrations in sea water. We will calibrate our system and we will estimate the analytical model that follows the behavior of our low cost turbidity sensor.**

Keywords-Turbidity; turbidity optical sensors; low cost sensor.

I. INTRODUCTION

Nowadays, it is important to have wide knowledge in underwater parameters because large changes can be important in some areas such as aquaculture, pollution monitoring, sewage treatment or oceanography.

The Water Framework Directive [1] proposes to measure some physicochemical parameters like temperature, dissolved oxygen or salinity, among others, to evaluate the quality of all water bodies (including marine water till one nautical mile from the coast).

In last years, the most common methods to analyze the water and to define these variables were sampling in-situ and the subsequent analysis at laboratories. The use of sensors to perform this kind of measurements is becoming increasingly common because these methods can reduce the analysis cost of the physicochemical parameters. The use of sensor nodes can be a way to eliminate the manual process of sampling because the measurements are performed automatically in the environment (river, lake, fish farm, etc.) by the sensor nodes. The use of sensors reduces costs avoiding going to the field for taking samples. When measurements and analysis are performed directly in the medium by using sensors, the sampling process is enhanced and we are sure that the number of samples is representative.

The lifetime of a physical sensor is longer than a chemical sensor. The chemical methods used at laboratories use reagents consumed during the analyses. Physical sensors also let us measure in real time. If we use several sensors forming a network along a river, it is possible to have measures along any type of environment such as a river bed or in the surrounding area of a marine fish farm. For these reasons, the use of sensor networks is growing day by day and now, it is easy to find many examples [2].

The turbidity is defined as the decrease of transparency of a water solution caused by the presence of dissolved or suspended particles. The size of particles is also important although sizes lower than 0.45µm do not affect to this parameter. These particles reflect, scatter and attenuate the light [3]. Turbidity measurements are important for environmental monitoring and management. In some cases, the sampling process is complicated and it may alter the environmental conditions. The alteration of a sample can make it not being a representative sample, so it should not be taken into account. Turbidity is expressed in Nephelometric Turbidity Units (NTU).

Turbidity values can vary by changes in the composition of the solids and some dissolved substances in water. In the seas, oceans and rivers those substances can have different origins and different effects. On one hand, the solids can come from the erosion of the emerged zones which provide nutrients to the water. On the other hand, the solids can come from different industrial effluents and in this case they can be dangerous for the environment. The suspended solids can also have a biological origin. They can be different type of microorganisms like phytoplankton, zooplankton or organic particles matter. Consequently, turbidity can indicate a wide range of situations (pollution, eutrophication or the increase of solids in the water mass). The sediments in suspension can cause several environmental damages such as benthic smothering, irritation of fish gills and the transport of absorbed contaminants. These sediments also produce an attenuation of the light penetration and this affects to the photosynthesis process [4]. The light dispersion can be used to measure the turbidity at specific wavelength and at specific angle, usually 90º [3].

In this paper, we propose the design of a low cost turbidity sensor. In addition, we have performed a set of measurements in sea water for several turbidity levels. Finally, we have performed a verification process to ensure that our system is working correctly.

The rest of this paper is structured as follows. Section 2 shows a review of the main methods which are used to measure the turbidity. Our low cost turbidity sensor is presented in Section 3. Section 4 explains the processes carried out to measure and calibrate our system. Finally, Section 5 shows our conclusion, compares the prize of our proposal with existing proposals and presents our future work.

 ISBN: 978-1-61208-296-7

II. RELATED WORK

Turbidity can be directly and quantitatively measured using several methods. In this Section, we are going to present the main ones.

A. Secchi disc and variations

Steel and Neuhausser [5] performed a study in Skagit River (Western of Washington State) where the turbidity was measured with 3 discs: horizontal black disc, horizontal Secchi disc and vertical Secchi disc. They also measured the clarity of the river with a nephelometer. Finally, they compared the results obtained by the nephelometer and the distances of the disc. These results are shown in Table I.

Larson et al. [6] addressed the possibility of reducing sources of error in measuring lake water clarity using a Secchi disk by investigating the use of beam transmissometer readings for estimating Secchi disk clarity. To do this, the authors investigated how Secchi disk works and performed the measurements in the Crater Lake in Crater Lake National Park, Oregon, USA. From their results, they extracted a prediction model to determine the water turbidity. The results demonstrated that this prediction model reliably estimated Secchi disk depths and could be used to significantly expand optical observations in an environment where the conditions for standardized Secchi disk deployments are limited.

B. Optical sensors

An optical turbidimeter bases the turbidity measure on the light detection at different angles. The name of the instruments changes according to the angle used to perform the measurements. If the angle is 90º, it is called nephelometer. If the angle is 0º, it is called absorbimeters. Because the turbidity is a value which depends on the quantity of suspended solids in water, the solids intercept the line of light causing scattering and absorption, so not all light arrives to the detector [7].

When the turbidity is low, the light scattered away from its original direction is lower. This is one of the most important problems when we try to measure low turbidity levels with a turbidimeter. To improve these measures, A. F. B. Omar and M. Z. B. MatJafri proposed the use of light detector in different angles, from 90 to 180º. This is called backscattering [7]. Other problems that optical sensors can have are the fouling, the high variability of the size and the reflectivity of particles [8].

Laser sensors are being investigated as an alternative method to the nephelometers. T. S. Melis et al. [9] developed a study of turbidity in Colorado River. They used Laser In-Situ Scattering and Transmissometry to measure the sand and thinner material. They measured the sensor output voltage in mV and the concentration of the sediments in mg/l. Their results showed the correlation in (1):

$$Sensor\ Output(mV) = [sediments]\left(\frac{mg}{l}\right) \cdot 0.971 - 1.942 \quad (1)$$

The interval of sediments concentration measured was from 10mg/l to 160 mg/L, and the recorded voltages were from 10 mV to 150 mV. Their results show that LISST-100-B is suitable for providing data of suspended solids of the Colorado's River, The measurement results presented a high confidence bound. It was around 95%.

Gray et al. in 2003 [10] took continuous in-situ turbidity measurements with the purpose of estimating the sediment concentrations in the Kansas River. They found a correlation of R=0.93, and the formula that relates the suspended sediment concentration (SSC) with the water turbidity was (2):

$$SSC\left(\frac{mg}{l}\right) = 1.797 * NTU * 0.905 \quad (2)$$

The range of values of their experience was between 10mg/l to 4000mg/l and between 10 NTU to 4000 NTU.

C. Acoustic sensors

Other method to measure the turbidity is the use of an Acoustic Dolpper Velocimeter (ADV) with backscattering. Chanson et al. [11] performed an experiment where the correlation between acoustic backscatter intensity and the sediment concentration showed a monotonic increase instead of the linear relationship. It is similar than the one that we can see between turbidity and suspended sediments [11]. They prepared some experiences in the laboratory and other ones in the estuarine zone of Eprapah Creek with low sediments concentration (Eastern Australia). Their results showed that the quality of water and the composition of sediments affect to the calibration curves. For this reason, each sensor must be calibrated with the water and soil material of the study area.

D. Other methods

There are several papers where authors use satellite images (LANDSAT) to measure the turbidity in water bodies [12]. However, this methodology is not interesting for our purposes because we need a continuous measurement process. In addition, we want to have values in a short period of time. Satellite systems cannot provide us short temporal scale (every hour), enough spatial resolution (few meters) and low economic prices.

TABLE I. MEASURES AND CORRELATION BETWEEN DISC AND NEPHELOMETER

Type of disc	**Results**			
	Authors	***Depth***	***NTU***	***Value of r***
Horizontal black disc	E. A. Steel and S. Neuhausser [5]	3.5 – 5 (cm)	1 - 4	r =-0.86
Horizontal Secchi disc	E. A. Steel and S. Neuhausser [5]	4 – 5 (cm)	1 - 4	r = -0.85
Vertical Secchi disc	E. A. Steel and S. Neuhausser [5]	3 – 5.5 (cm)	1 - 4	r =-0.86

 ISBN: 978-1-61208-296-7

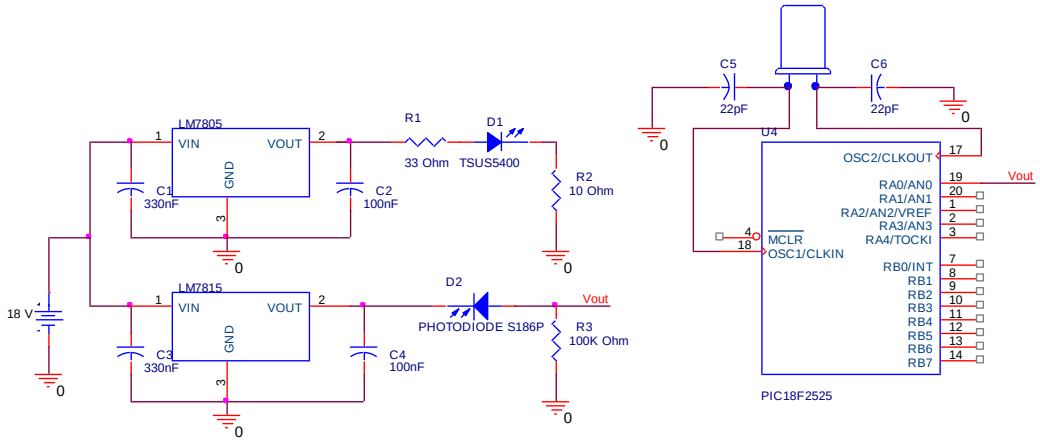


Figure 1. Esquematic of our low cost turbidity sensor.

The pressure-differential principle can also be used to measure the turbidity. This method is promising to be used in highly concentrated stream flows [10]. But, we cannot use this method because our measures are performed in the sea where the concentration of sediments is usually low.

III. LOW COST TURBIDITY SENSOR PROPOSAL

In this section, we are going to present our low cost turbidity sensor. This section also shows its fabrication cost.

A. *Our proposal*

Our system is based on the use of an infrared LED as a source of emission and a photodiode as a receiver. Both elements are disposed at an angle of 180° and they are faced so that the photodiode can capture maximum infrared light from the LED. The infrared LED and the photodiode are placed at a distance of 4 cm.

TSUS5400 is an infrared emitting diode using GaAs technology molded in a blue-gray tinted plastic package. This infrared diode has a peak wavelength of 950 nm and its angle of half intensity is 22°.

S186P is a high speed and high sensitive photodiode in a plastic package. It is an IR filter, spectrally matched to GaAs or GaAs on GaAlAs IR emitters (≥ 900 nm). The large active area combined with a flat case gives a high sensitivity at a wide viewing angle. The angle of half sensitivity is ±65º. S186P is covered by a plastic case with IR filter (950 mm) and it is suitable for near infrared radiation.

The transmitter circuit is powered by a voltage of 5V while the photodiode is supplied with a voltage of 15V. To achieve these values, we use two voltage regulators of LM78XX series, where XX is 05, for the transmitter circuit, and 15 for the receiver circuit. This type of regulatory permits a maximum output current of 1 A.

In our system, the water turbidity is proportional to the potential difference registered in the resistor R3. To process the voltage values recorded in the receiver circuit a microcontroller is used. This element is responsible for taking the data and performs the necessary calculations to estimate the turbidity value of the analyzed solution. The model we have used is the PIC18F2525 which is fed at 5V. We can use the microcontroller to display information via a LCD display or to send the information through the RS232 port to a personal computer.

Fig. 1 shows the schematic of our turbidity sensor.

B. *Cost of our turbidity sensor*

The electronic components used in our turbidity sensor can be easily found. We have asked for their prize in a Spanish electronic distributor. The prizes of the main components are shown in Table II.

TABLE II. PRIZE OF MAIN COMPONENTS

Component	Prize of components (in €)
Voltage regulator LM7805	0.80
Voltage regulator LM7815	0.86
IR LED TSUS5400	0.176
Photodiode S186P	0.94
PIC 18F2525	5.71
All Capacitors	2
oscilator	2.18

As Table II shows, the prize of the needed electronic components is very low

IV. TEST BENCH AND MEASUREMENT RESULTS

Our main problem when taking the turbidity measurements was to determine the turbidity level of each solution test probe. These values were necessary in order to estimate the analytical model of our system. This section explains the process carried out to take the measurements and the calibration process of our system.

A. *Used elements*

On the one hand, we employed a calibrated turbidimeter to know the accurate value of turbidity of each test probe to be measured. The used device is Turbidimeter Hach 2100N. This turbidimeter can take measurements in two modes (ratio

 ISBN: 978-1-61208-296-7

mode and non-ratio mode). We have used the non-ratio mode, which measures the turbidity by using a detector placed at 90 degrees.

All measures have been performed for seawater. The salinity of the sea water depends on the geographic position. Our samples were taken from the east coast of the Mediterranean Sea (Spain). The average water salinity ranges from 36 to 38 grams per liter. pH is 8.07 and its temperature is 21.1ºC.

Regarding to the elements used to generate the dissolutions with different turbidities, we have used a mixture of clay and silt. The silt is a loose material with a grain size between fine sand and clay. It is an incoherent classic sediment transported in suspension by rivers. The particle size of our sample is between 0.002 mm and 0.06 mm.

The glass containers used in our test had a volume of 30ml. First, took four calibration samples in the turbidimeter Hach 2100N. Then, we took 10 samples with turbidities. Next, we performed more samples at lower concentrations than at higher ones because lower concentrations are most usual. The concentration samples we prepared are shown in Table III. We have selected samples with low turbidity because for devices low turbidities are more difficult to measure these.

Fig. 2 shows the relationship between the amount of solute and the concentration of the test probe.

We can extract the relationship between both values using (3) with a correlation coefficient of 0.9999.

$$C = 33.333 \cdot S + 10^{-13} \quad (3)$$

where C represents the concentration of the dissolution in mg/l and S represents the amount of solute in g. Because the independent term is extremely small value, we can approximate this relationship as (4):

$$C = 33.333 \cdot S \quad (4)$$

Each prepared sample has a turbidity value. Fig. 3 shows the relationship between the concentration of silt-clay in ocean water and the water turbidity value.

Equation (5) shows the relationship between both parameters with a correlation coefficient of 0.9999.

$$T = 0.517 \cdot C + 3.0027 \quad (5)$$

where, T represents the turbidity in NTU and C is the dissolution concentration in mg/l.

Finally, we can express the turbidity as a function of the amount of clay-silt dissolved combining (4) and (5). It is given by (6).

$$T = 17233.161 \cdot S + 3.002 \quad (6)$$

where, T represents the turbidity in NTU and S, the amount of solute in g.

B. *Obtained results*

After measuring the turbidity of each sample, we placed each one in our system in order to measure their voltage values.

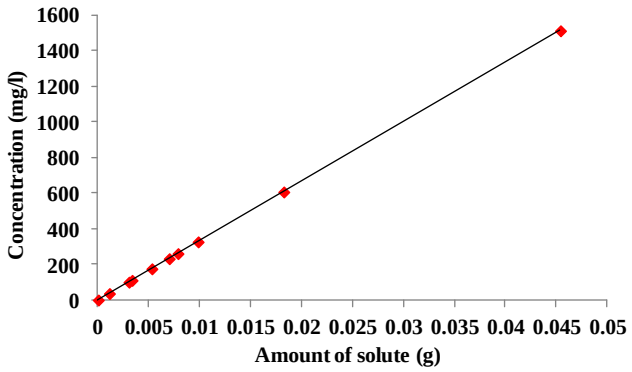


Figure 2. Value of dissolution concentration as a function of the amount of solute.

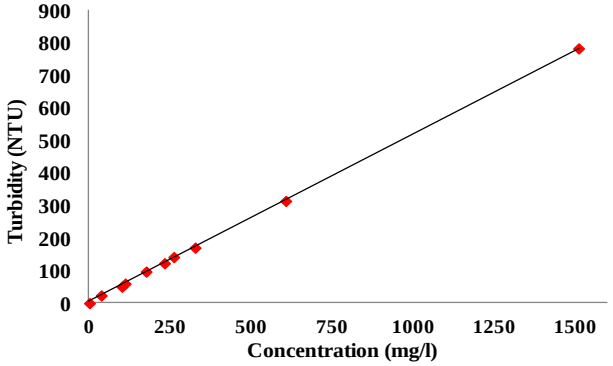


Figure 3. Value of dissolution turbidity as a function of the dissolution concentration.

TABLE III. SAMPLES, THEIR CONCENTRATION AND TURBIDITY

Sample	Samples and Their Concentration									
	1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*
Amount of clay and silt added (g)	0	0.00109	0.00301	0.00331	0.00526	0.00696	0.00782	0.0098	0.01821	0.0453
Concentration (mg/l)	0	36.33	100.33	110.44	175.33	232.1	260.66	326.66	607	1512.59
Turbidity (NTU)	0.072	23.7	50.8	60.1	97.2	123	142	171	385	785

 ISBN: 978-1-61208-296-7

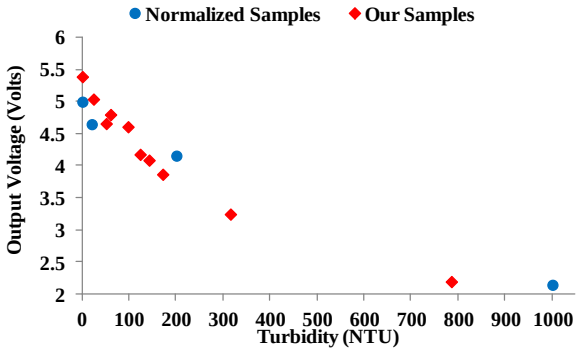


Figure 4. Output voltage as a function of the turbidity value.

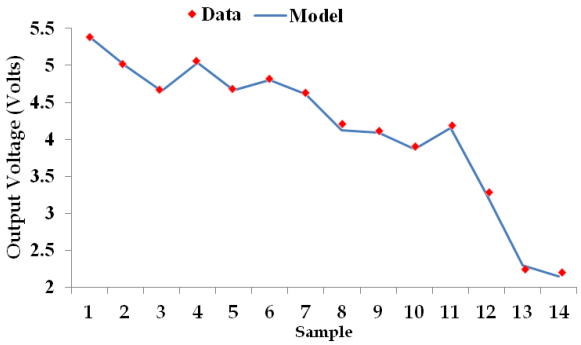


Figure 5. Gathered Data vs. Model response

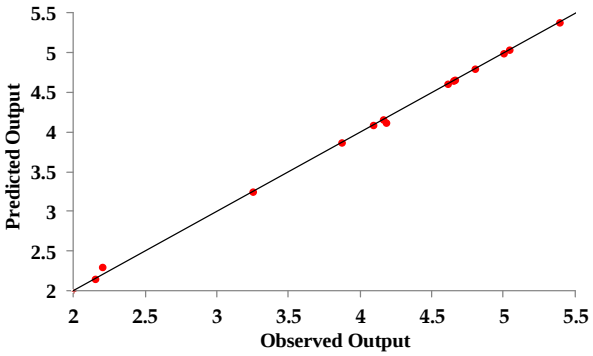


Figure 6. Observed output vs. Predicted output

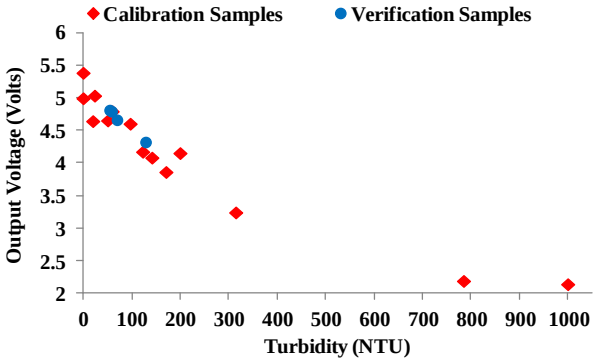


Figure 7. Output voltage value for the verification samples

$$y = a + \frac{b}{(c+x+d\cdot\sin(e+f\cdot x)+g*\cos(h+i\cdot\sin(j+k\cdot x)))} \quad (7)$$

$$v = 0.9767 + \frac{1550}{(353.4+t+31.69\cdot\sin(5.435+2.004\cdot t)+19.34*\cos(6.141+18.85\cdot\sin(5.443+2.004\cdot t)))} \quad (8)$$

Fig. 4 shows the gathered output voltage value for each the turbidity. The prepared samples are shown in red while the standard samples are shown in blue. We can see that the behavior of our system is not linear. Its behavior can be approximated to a linear function up turbidity values of 300 NTU. But, for higher turbidity values, the behavior is closer to a curve function with an error lower than ±1%.

We used Eureqa Formulize [13] to estimate the analytical model. Using these 14 samples, we have estimated the analytical expression that models the behavior of our turbidity sensor. The program provided us several models but the one which provided best results is the model show in (7). Equation (8) relates the turbidity with the output voltage of our system, where, *t* represents the turbidity in NTU and *v* the output voltage.

Fig. 5 shows the obtained values for each sample in blue (estimated, using (8)). The obtained output voltage values for the same samples are shown in red. As Fig.5 shows both values are close.

Finally, in order to check the accuracy of our system, we represent the output voltage value predicted by our model versus the value of output voltage obtained by our system. Fig. 6 shows this relation. As Fig.6 shows, most of the points remain in the black line, which indicates a perfect match.

From Eq. 8, Fig. 5 and Fig. 6, we observe that our equation has a correlation coefficient of 0.99897 and its maximum error is 0.1 Volts.

C. *Verification*

In order to test the accuracy of our system, we have compared the values obtained by the turbidimeter Hach 2100N and the values obtained by our turbidity sensor.

In the verification test, we have chosen low concentration samples, which means low turbidity. In these situations, it is more difficult to measure accurately the turbidity. Commercial Turbidimeters have also problems in low turbidities. In order to perform this step, we have prepared four samples with unknown concentrations. The output voltages obtained for each sample are shown in table IV. Estimating the turbidity values by using (8), we have introduced our samples (in blue) in (7). It shows the behavior of our system. In order to compare the estimated turbidity values with the measured values and the error for the four samples, we provide Table V.

TABLE IV. UNKNOWN SAMPLES

Samples	**Samples to verify**			
	1	***2***	***3***	***4***
Output Votlage	4.67 V	4.82 V	4.80 V	4.33 V

 ISBN: 978-1-61208-296-7

TABLE V. VERIFICATION RESULTS

Results	Samples to verify			
	1	*2*	*3*	*4*
Turbidity obtained	73.66	55.36	57.80	115.2
Turbidity (Hach 2100N)	70	54.8	59	129
Error in Turbidity (%)	5.22	1.03	2.03	10.75

The biggest error in turbidity is registered for the fourth sample (10.75 %) meanwhile the samples with lower turbidity present an error of 1.03 %.

V. CONCLUSION

In this paper, we have presented the circuit of a low cost turbidity sensor. In addition, we have performed several tests in sea water to model and to calibrate the operation of our turbidity sensor.

Although sea water usually presents low turbidity levels and we have focused our system in lower values; we have also added some values of higher turbidity level. For low turbidity values, we have obtained a maximum error of 5.22%. We have also compared the prize of our device with commercial ones in Table VI.

TABLE VI. COMMERCIAL TURBIDIMETERS

Model	Commercial turbidimeters		
	Manufacturer	*NTU range*	*Prize*
PCE TUM 20	PCE Holding GmbH	0 - 1000	285 €
HI 88713-02	Hanna Instruments Deutschland GmbH	0 - 4000	1,287 €
HI 98713-02	Hanna Instruments Deutschland GmbH	0 - 1000	924.75 €
Turbiquant® 1100 IR	Merck Millipore	0 - 1000	1,570 €
Turbiquant® 3000 IR	Merck Millipore	0 - 10000	4.240 €

Seeing these prizes, we think we are able to develop a low cost turbidity sensor.

In our future work, we will improve our system. First, we believe that there are some external factors such as the measurement system disposal and the electronic elements placement, significantly affect the sampling process. So, we will improve our measurement system.

On the other hand, we will improve the way to find the analytical relationship between the measured parameters by adding more mathematical operations. Additionally, we would like to perform other calibrations with others waters.

After these improvements, we would like to test our sensor in marine fish farms [14].

ACKNOWLEDGMENT

This work has been partially supported by the "Ministerio de Ciencia e Innovación", through the "Plan Nacional de I+D+i 2008–2011" in the "Subprograma de Proyectos de Investigación Fundamental", project TEC2011-27516, and by the Polytechnic University of Valencia, though the PAID-05-12 multidisciplinary projects, Ref: SP20120420.

REFERENCES

[1] Directive 2000/60/EC of the European Parliament and of the Council of 23 October 2000 establishing a framework for Community action in the field of water policy. Official Journal of the European Communities. Available at: http://faolex.fao.org/cgi-bin/faolex.exe?rec_id=017501&database=FAOLEX&search_type=link&table=result&lang=eng&format_name=@ERALL (Last access: April 22, 2013).

[2] M Garcia, D Bri, S Sendra, and J Lloret, Practical deployments of wireless sensor networks: A survey, International Journal on Advances in Networks and Services vol. 3, no 1 & 2, 2010, pp. 170-185.

[3] Directive 2001/42/EC of the European Parliament and of the Council of 27 June 2001 on the assessment of the effects of certain plans and programmes on the environment. Official Journal of the European Communities. Available at: http://faolex.fao.org/docs/pdf/eur34996.pdf (Last access: June 10, 2013).

[4] A. C. Ziegler, "Issues related to use of turbidity measurements as a surrogate for suspended sediment". Turbidity and Other Sediment Surrogates Workshop, Reno, Nevada, USA, April 30 – May 2, 2002.

[5] E. A. Steel and S. Neuhausser, "A comparison of methods for measuring visual water clarity", Journal of the North American Benthological Society; vol. 21, no. 2, 2002, pp. 326-335

[6] G. L. Larson, R. L. Hoffman, B.R. Hargreaves, and R. W.Collier, "Predicting Secchi disk depth from average beam attenuation in a deep, ultra-clear lake". Hydrobiologia, vol. 574, no. 1, 2007, pp. 141-148.

[7] A. F. Bin Omar and M. Z. Bin MatJafri, "Turbidimeter design and analysis: a review on optical fiber sensors for the measurement of water turbidity", Sensors; vol. 9, no 10, 2009, pp. 8311-8335.

[8] D. H. Schoellhamer and S. A. Wright "Continuous measurement of suspended-sediment discharge in rivers by use of optical backscatterance sensors" Workshop on Erosion and Sediment Transport Measurement in Rivers: Technological and Methodological Advances, Oslo, Norway, June 19-21, 2002, pp. 28-36

[9] T. S. Melis, D. J. Topping, and D. M. Rubin, "Testing laser-based sensors for continuous in situ monitoring of suspended sediment in the Colorado River, Arizona", Workshop on Erosion and Sediment Transport Measurement in Rivers: Technological and Methodological Advances, Oslo, Norway, 19-21 June 2002. Pp.21-27.

[10] J. R. Gray, "US Geological Survey research on surrogate measurements for suspended sediment" 1st Interagency Conference on Research in Watersheds Benson, Arizona, USA, Oct. 27-30, 2003, pp. 95-100.

[11] H. Chanson, M. Takeuchi, and M. Trevethan "Using turbidity and acoustic backscatter intensity as surrogate measures of suspended sediment concentration in a small subtropical estuary" Journal of Environmental Management, vol. 88, no 4, 2008, pp. 1406-1416.

[12] R. Li and J. Li,"Satellite remote sensing technology for lake water clarity monitoring: an overview". Environmental Informatics Archives, Vol.2, 2004, pp. 893-901.

[13] Eureqa Formulize web site. Available at: http://formulize.nutonian.com (Last access: April 22, 2013).

[14] M. Garcia, S. Sendra, G. Lloret, and J. Lloret, "Monitoring and control sensor system for fish feeding in marine fish farms". IET Communications; vol. 5, no. 12, 2011, pp. 1682-1690.

 ISBN: 978-1-61208-296-7

Two New Sensors Based on the Changes of the Electromagnetic Field to Measure the Water Conductivity

Lorena Parra, Vicente Ortuño, Sandra Sendra, Jaime Lloret
Integrated Management Coastal Research Institute (IGIC)
Universidad Politécnica de Valencia, Spain
loparbo@upvnet.upv.es, vortunyo@eln.upv.es, sansenco@posgrado.upv.es, jlloret@dcom.upv.es

Abstract— **Water salinity and conductivity values are always required when measurements from the sea water are needed (e.g., for water pollution, in marine fish farms, feeding industry, etc.). They can be measured directly or indirectly. In order to select the best option, three main issues should be taken into account. The first one is the periodic need of calibration because of the system wear, the second one is the cost of the components of the system, which could make to deploy an expensive device, and the third one is the accuracy of the deployed system. In this paper, we propose two cheap conductivity sensors that do not need periodic calibration. The first one of them is composed by a solenoid and a commercial magnetic field sensor that detects the changes in the magnetic field when different materials are introduced in the center of solenoid. The second one is composed by two overlapped copper coils that, through one of them, are circulating energy, and induce in the other one. The different environments of the coils produce changes in the induced voltage in the second coil.**

Keywords-electrical conductivity; solenoid; copper coils; magnetic field; salinity.

I. INTRODUCTION

The conductivity is defined as the capability of a matter or medium to permit the pass of electricity through it. It is measured in Siemens/meter or microhoms/cm [1]. Electric conductivity is also defined as the natural property of each body which represents how ease the electrons can pass through it. The electrical conductivity of liquids is related with the presence of salts, which generates positive and negative ions, because they are able to carry electric energy through the solution. Those ions are called electrolytes or electrolytic conductors. The electrical conductivity depends on the temperature of the solution. Because the temperature changes, it can also change the values of the ions, solubility, and solution viscosity among other issues [2].

There are different ways to measure the electrical conductivity of the water. The traditional one is to measure the conductivity or resistance offered by the water. It can also be measured by using diamagnetic and paramagnetic proprieties of the water with different concentration of salty ions. The paramagnetic substances increase the value of the magnetic field. Moreover, the diamagnetic substances drive down the magnetic field. Generally, each material has both kinds of behaviors, but predominates one of them.

The magnetic fields are composed of electrical charges, which react with the environment. Those charges can attract or repel themselves and their behavior depends on the chemical or physical forces of the environment. Electric and magnetic charges represent different aspects of the same event. When there is no electric or magnetic charge, the electron's loads are not agitated. When an attractive force is applied, the electron's loads are agitated and begin to move in the direction of the applied force. In the case of water, its chemical composition determines the effects in the magnetic field. The interaction occurred between the electrical charges and the water molecules can cause that some atoms lose their electrons, those atoms are ionized or charged. As a result, these atoms attempt to recover the missing electrons. The combination of the ionized atoms and the magnetic fields causes the formation of an electric current in the water.

When an electromagnetic field pass through a material or medium, the measurement of the changes of the electromagnetic field can bring information about some of its proprieties. In the water medium, the measurement of electric conductivity can bring information about water quality and the quantity of dissolved salts. This is important in many areas such as water management, agriculture, aquaculture or groundwater supplies. In the case of agriculture, it is very important to know the salinity of the water used in the irrigation process; because when the soil is irrigated with water that contains high concentration of salts can produce salinization of the irrigated soils in the long term. It is estimated that 50% of cultivated fields are suffering this kind of salinization [3]. In the case of aquaculture in fresh or sea waters, changes in salinity can cause the death of the cultured species, causing huge economic losses. Moreover, saline intrusion is causing great damages in the groundwater supplies, which would lead to obtain not drinkable waters with the time. Unpolluted fresh water is becoming a limiting resource in some regions, so the saline intrusion in the aquifers of those regions must be controlled to ensure the availability of the water quality.

All these problems can be prevented and corrected using the proper control. Sensor networks, where sensor nodes are sensitive to conductivity changes, can bring an early warning signal, which allows applying the necessary measures to prevent harmful effects. In order to develop this sensor network, the first step is to develop a physical sensor able to measure the conductivity, which must be as cheap as possible, because, for example, to measure the environment of an aquaculture installation many sensors are needed. Low maintenance is also required for the sensors, so the contact between the water and the sensor should be minimized.

 ISBN: 978-1-61208-296-7

If we only want to detect big changes in salinity, for example, when an object is changed from fresh water to sea water, there is no needed to have high sensibility. Moreover, in some cases, the most important issue is to reduce the maintenance of the sensor as much as possible and try to eliminate the need for periodic calibration. The goal in these cases is to place the sensor during long periods of time.

The aim of this paper is to develop two new conductivity sensors but with the purpose of having low manufacturing costs and low maintenance cost, which can be applied to different areas. We will create a magnetic field and make it pass through the water with different conductivities. It will allow us to detect if the changes in the magnetic field are correlated with the changes of the electrical conductivity. In order to measure the changes in the magnetic field we developed two different sets of assays. In the first one, we create a magnetic field with a solenoid and measure the values of that field with a commercial magnetic field sensor. In the second one, we use two copper coils, one of them has an electrical current and induces this current over the other coil, the measure of the voltage of the second coil gives information about the conductivity.

The rest of this paper is structured as follows. In Section 2, we review other articles developing conductivity sensors. In Section 3, the structure of the both sets of laboratory assays is described. Section 4, describes the assays made at laboratory to evaluate the electrical conductivity detection when there are changes in the magnetic field. Section 5 shows the obtained results. In Section 6, we make a comparison of the prices of commercial sensors and the developed sensors. Finally, Section 7 presents our conclusion and future work.

II. RELATED WORK

There are some works in the related literature where the authors developed conductivity meters. This section presents a review of those works.

In 2007, Medrano et al. [4] developed their own conductivity meter for liquids with low electrical conductivity (measuring directly the conductivity of the water). The minimum value that they were able to measure was 200pSm−1, with an error of 10%. They measure with different distances between both electrodes (0.5 to 2.5 mm) and different voltages (-10 to 10V).

Wei at al. [5] proposed in 2010 a new seawater conductivity sensor (also based on the capacity of water to transmit the electricity), which uses a bipolar pulse to avoid the effect of electrode polarization. They also propose a temperature composition with different formulas for 3 ranges of temperature 1 ºC to 10 ºC, 10 ºC to 20 ºC and 20 ºC to 30 ºC. The sensor is able to self- compensate and self-tuning.

Ramos et al. [6] created a low cost in-situ four electrode conductivity cells in 2006. It is suitable to take measurements from estuarine waters.

The other way to measure the conductivity through the alteration of the magnetic field is used only in the study of saline soils [7, 8]. In those works, two coils are used, the energy passes though one of them and a charge to the other is induced. The charge on the second coil depends on the salinity of the soil.

As far as we know, the use of the interaction in the electromagnetic field has not been used to measure the water conductivity yet (at least in the published works). Moreover, there are no papers describing the process that occurs when an electromagnetic field passes through water with different electrical conductivities. However, few commercial sensors, that use two coils to measure the induction, exist. In this paper we pretend to identify the effects of different electric water conductivities when an electromagnetic field generated by a copper solenoid passes through it.

III. STRUCTURE OF CONDUCTIVITY SENSOR

Our purpose is to detect changes of the electromagnetic field and relate it with the conductivity of the water where it is passing through. It will let us create an electrical conductivity detector. In order to measure the electrical conductivity values, we used a commercial sensor: CM 35 + [a]. By using two different methods, we developed two different inductive conductivity sensors.

For the first method, we prepared an assembly with a solenoid without core that generates the electromagnetic field. In the center of the solenoid, we introduced an assay tube, which is used as container for the water samples. Moreover, in the center of the solenoid (where the magnetic field is higher), inside the assay tube, we inserted a magnetic field sensor. The solenoid was powered by a Direct Current (DC) generator or connected to an alternative current trough a transformer from 220V to 12V, depending on the desired output measurements (in Direct Current or Alternating Current, AC); see the explained in Figure 1.

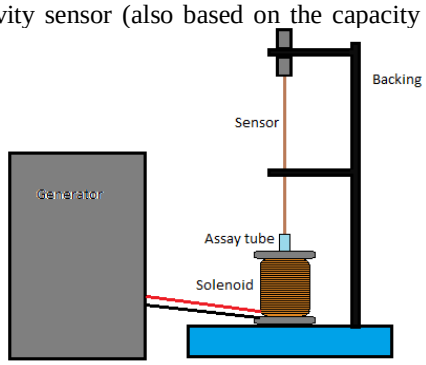


Figure 1. Structure of the first set of laboratory assay.

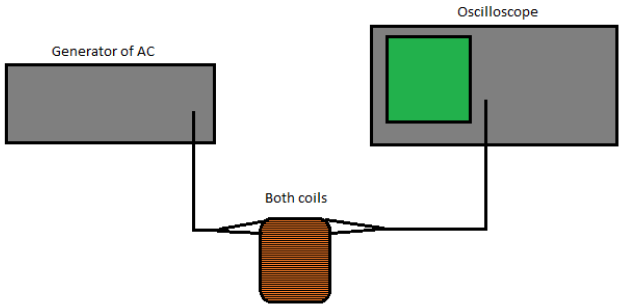


Figure 2. Structure of the second set of laboratory assay.

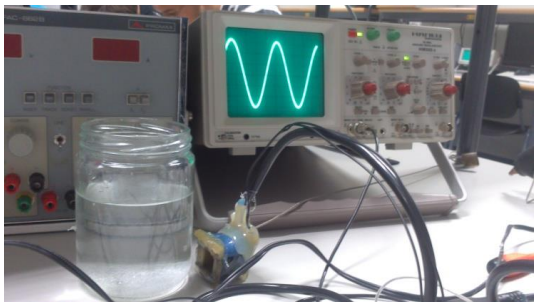


Figure 3. Laboratory set-up.

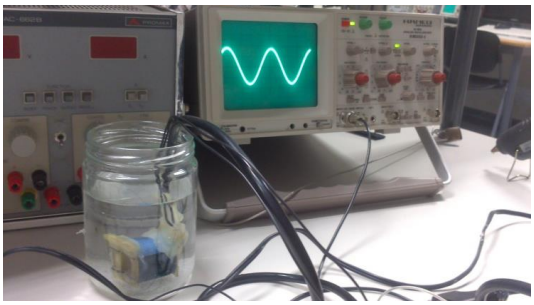


Figure 4. Laboratory set-up with the probe inside the water.

The whole structure was fixed using a laboratory support to assure that when the solenoid and the assay tube are removed (in order to change the solution), the position of the sensor respect to the solenoid is maintained. This is very important, because when the position of the sensor over the centre of the solenoid changes, vary the values of magnetic field. This backing system is useful only for the height and the vertical movement. The horizontal movement has less importance because the width of the assay tube, the centre of the solenoid and the sensor are almost the same and that precluded the horizontal movement. We built a second solenoid without the assay tube in order to avoid the potential interferences caused by the glass. In this solenoid the wire involves a plastic tube that contains the water samples. The sensor used in this set of assays was the Hall Axial Payme probe. It was connected to a Tesalameter Phaywe with a range of measurement from 0 to 2000mT.

We also developed a second experiment. We used two cooper coils. In this case, the wire has small diameter. Those coils are overlapped and do not have a core. The turn relation is 1:36.66. In one coil we introduced AC, then, we measured the induced current in the other coil. The explained structure is show in Figure 2.

Figure 3 shows the probe outside the water. The generator sends a sine signal through the coil. The coil core is air. The oscilloscope shows the obtained result. Figure 4 shows the probe inside the sea water. We can see in the oscilloscope that different results are obtained.

IV. LABORATORY ASSAY

We have done two sets of analysis. The first one uses the solenoid and the commercial sensor. The purpose of these assays was to demonstrate that differences in the environment can produce alterations in the magnetic field. The second set of analysis was aimed to obtain more sensibility and the reduction of the cost of the sensor.

A- First set of assays: Solenoid + Commercial sensor

The initial assays were done connecting the solenoid to the DC generator. We prepared 5 solutions with different ntration of salts; consequently, they present different conductivity. The expected behaviour was to vary the magnetic field when we place probes with different water conductivity.

In order to find the best voltage to take measurements and obtain the correlation between conductivity and magnetic field values, we prepared a scanning. We measured all the samples with different voltages, those voltage and current generated in the solenoid are shown in Table I.

TABLE I. VALUES OF VOLTAGE AND CURRENT REGISTERED IN THE FIRST SOLENOID

Voltage (V)	5	10	15
Current (A)	0.2	0.4	0.6

TABLE II. MAGNETIC FIELD (MT) VALUES WHEN MEASURING DIFFERENT CONDUCTIVITIES AT DIFFERENT FREQUENCIES WITH 12V (AC). SOLENOID 1 (WITH THE ASSAY TUBE).

Conductivity (mS)	**15Hz**	**150Hz**	**1500Hz**	**15000Hz**
Without assay tube	269	198	26	3
0,002	269	198	26	3
213	296	198	26	3

TABLE III. MAGNETIC FIELD (MT) VALUES WHEN MEASURING DIFFERENT CONDUCTIVITIES AT DIFFERENT FREQUENCIES WITH 12V (AC). SOLENOID 2 (WITHOUT THE ASSAY TUBE.)

Conductivity (mS)	**15Hz**	**150Hz**	**1500Hz**	**15000Hz**
Without water	0.17	0.18	0.16	0.03
0,002	0.16	0.18	0.15	0.03
213	0.16	0.18	0.15	0.03

TABLE IV. MAGNETIC FIELD (MT) VALUES WHEN MEASURING DIFFERENT CONDUCTIVITIES AT DIFFERENT VOLTAGES WITH DC

Environment	**Measured at 5V**		**Measured at 10V**		**Measured at 15 V**	
	Test A (mT)	**Test B (mT)**	**Test A (mT)**	**Test B (mT)**	**Test A (mT)**	**Test B (mT)**
Air	2,89	2,90	8,66	8,72	14,89	14,51
Water with conductivity (mS)						
0,002	2,91	2,93	8,68	8,71	14,86	14,56
0,405	2,93	2,96	8,69	8,89	14,82	14,66
191,4	2,93	2,98	8,69	8,89	14,81	14,64
285	2,97	3,01	8,70	8,90	14,82	14,59
213	2,96	3,00	8,73	8,92	14,78	14,62

 ISBN: 978-1-61208-296-7

Later, we measured these samples with AC at 12V. With the same purpose as before, now we change the frequency of the current in order to find the best point to perform the rest of measures and obtain the most accurate correlation. The frequencies used to take measurements were 15Hz, 150Hz, 1500Hz and 15,000Hz. In this case, we used two solenoids, the first one was the same used in the DC, the solenoid with an assay tube inside, and the second one was the solenoid with the plastic tube. The purpose was to reduce the influence of the glass tube.

In order to complete the first set of assays, we took measurements with other water samples by using the first solenoid at higher DC voltages.

B- Second set of assays: Two coils

The second set of laboratory assays was performed using the two overlapped coils. The first one was connected to AC and the second one provided an electromagnetic induction. The core of those coils was removed, allowing us to use the air or water as the coil core. Then, we took measurements in both environments, air and water (in this second case we immersed both coils in a container full of water with high salinity). The measurement of the induced voltage was taken with an oscilloscope by measuring from peak to peak the sine wave.

In a first assay, we performed several changes of the working frequency. We observed the difference of the induced voltages in the second coil when the environment changes (air or water). The second assay was also performed with different frequencies, but in this case the voltage of the second coil was fixed to 2.8V in air. Then, we introduced inside a container with salty water and compare it with the obtained voltage after this change. The aim of this assay is to find the point where the change of the voltage is higher when we change the environment. Because this point depends on each coil, we made some tests to use it in future assays.

V. RESULTS

A- First set of assays: Solenoid + Commercial sensor

First, we describe the results obtained in the first set of analysis performed with the solenoid and the commercial sensor. We are going to start showing the results obtained with AC; see Table II. We have not appreciated any correlation between the magnetic field and the conductivity of the water at 12V at any frequency. Thus this method is not useful with the used equipment (maybe with a sensor with higher sensibility we will be able to take this measure).

We repeated the same assays with the second solenoid in order to know if the problems obtained in the previous test are caused by the influence of the glass tube. The results are show in Table III. In this case we have not also appreciated any correlation between the magnetic field and the conductivity of the water at 12V; but, we have seen a difference in the magnetic field when the environment changes (air or water) only at 15Hz and at 1500Hz.

Now we present the results obtained with the assays when using DC. In this case, we only take measurements with the first solenoid because the other solenoid have less turns so the generated electromagnetic field is lower. We observed that the influence of the glass tube is not so important.

First, we took measurements with 5 samples with different concentration of salt and different voltages (shown in Table I). The conductivity of the samples has 3 intervals, 1 with the lowest conductivity, 0.002mS, other with low conductivity, 0.405 mS, and 3 samples with high conductivity 192 mS, 213 mS, and 285mS. This distribution allows us to see different issues. First of all, it is possible to distinguish between very different conductivities (low values and high values). Second, it allows us to know the different sensitivities in high conductivities or low conductivities, because the sensitivity is different at different ranges for electrical conductivity. The values of the magnetic field detected inside each sample are shown in Table IV. In this case, we did two series of measures at each voltage, to know the repeatability. Those measures are called Test A and Test B, Test A was done before and when it was finished we did the measures of Test B.

We performed two tests in order to take values for different voltages. This showed us that the sensibility increases, but, at same time, it caused some problems, because a change in the position of the sensor in the magnetic field causes variations in the lecture of the value. At high voltage values, the error committed in a single value is higher than at low voltages. It is so important when taking measurements from different ranges.

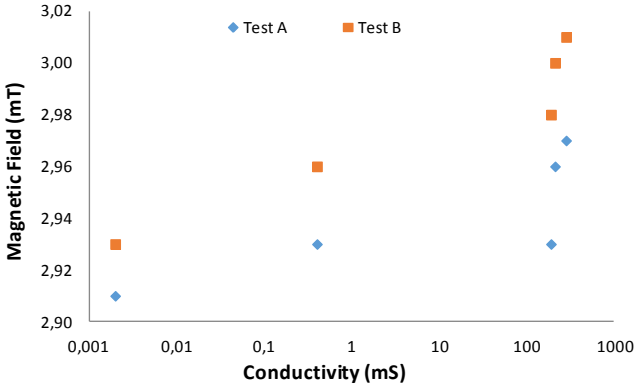


Figure 5. Representation of the data for 5V

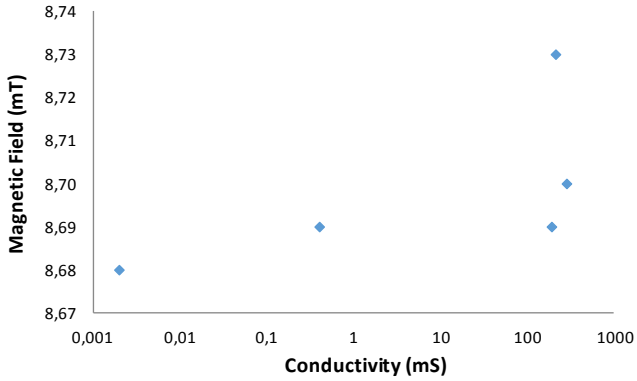


Figure 6. Representation of the data for 10 V for test A

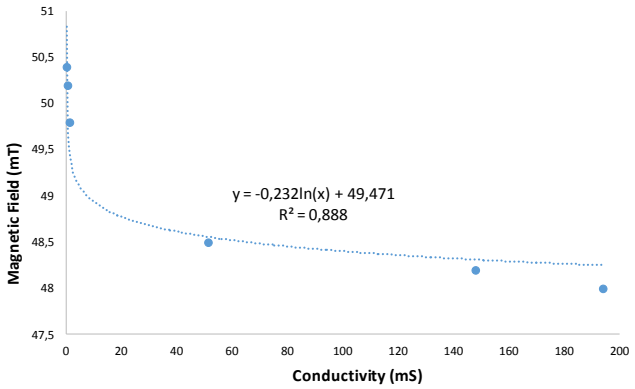


Figure 7. Representation of the data for 20V for test A

In Figure 5, we can see the measurements obtained at 5V. In this case, the results show that is possible to distinguish between different conductivities. At low conductivity, values the sensitivity is higher than at high conductivity values. Moreover, the magnetic field increases with the conductivity.

In Figure 6, the measures obtained at 10 V for test A are shown. The values are similar to the values obtained when measuring at 5V. It is possible to distinguish low values with high values, but not to distinguish between high values. At 10 V, in one case, the magnetic field value increases with the conductivity (test A) and in the other case decreases (test B).

We performed the test at 15 V two times, but both times yield confusing results, so they are not represented. We think that they may be caused by errors in the position of the sensor respect the magnetic field or because 15 V is not a good input voltage for the sensor.

We also performed measurements at 20V. The main problem was that at these voltages the coil starts to heat up, and this heat interferes with the measures. The results are show at Figure 7. In this case the values decrease when the conductivity increases and this follows (1). It defines the correlation between conductivity and the variation of the magnetic field.

$$\text{Magnetic Field } (mT) = -0{,}232 * \ln(\text{Conductivity } (mS)) + 49{,}471 \quad (1)$$

At different voltages, the changes of the magnetic field increase or decrease. We have observed that the water conductivity is different because there are differences between high and low conductivities so we introduce intermediate conductivities.

TABLE V. INDUCED COIL VOLTAGE AT DIFFERENT FREQUENCIES WHEN THE ENVIRONMENT CHANGES

Frequency (KHz)	Voltage (V) measured in	
	Air	Salty water
0,5	15	3
1	18	6,4
2	18	14
4	20	17,5
5	20	13,6
6	20	11,9
7	20	10,4
10	18	8
20	15	5
100	7,6	0,6

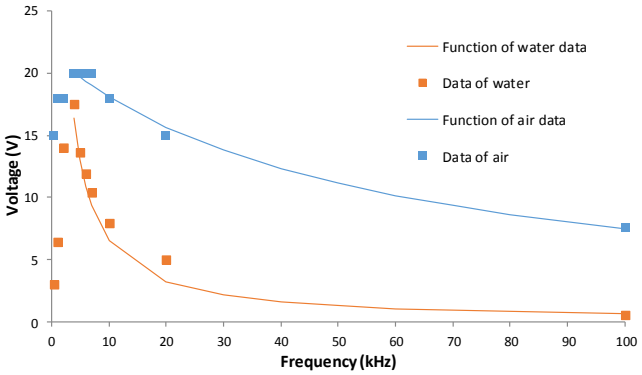


Figure 8. Representation of the data of Table V

B- *Second set of assays: Two coils*

In the first assay, we changed the frequency and we observed the voltage in the induced coil in both environments: water and in air. The results are show in Table V. They are represented in Figure 8. We observed that in the air, voltages increase until 4 kHz. Then, it is maintained until 7 kHz, where it starts to decrease. Otherwise, in water it increases until 4 kHz, where it starts to decrease. Both environments have the same behaviour with different peaks. These behaviours can be interpolated with the following (2) for air and (3) for water:

$$\text{Voltage } (V) = 23/(-1{,}071 - 0{,}01984 * \text{Frequency } (kHz)) \quad (2)$$
$$R^2 = 0{,}86$$
$$\text{Voltage } (V) = 653{,}4/\text{Frequency } (kHz) \quad (3)$$
$$R^2 = 0{,}86$$

These results demonstrate us that it is possible to distinguish between salty water and air at any frequency, but some frequencies have higher differences.

We need to know in which frequency the difference between environments in higher, and this is what we did in the second assay. We changed the frequency, but maintaining the voltage of the induced coil in air at 2.8 V. These results are show in Table VI and represented in Figure 9.

In Figure 9, we can see that depending on the frequency, the effect of the change of environment can have different sign (positive sign or negative sign). So the voltage in the water can increase or decrease with respect to the voltages in the air. The value of the voltage in the induced coil was allways 2,8V. The point where the sign changes is at 248kHz. From 10kHz to 248 kHz the change of environment (air to water) makes to decrease the voltage. This change is higher at 10kHz and decreases until 248kHz where is null. From 248kHz the change of the environment makes to increase the voltage and the difference of voltage increases when the frequency increases. Part of this data (from 10 to 1000kHz have a logarithmic behaviour and follow (4):

$$V_{\text{of induced coil}}(V) = 0{,}5133\ln(\text{Freq. (kHz)}) + 0{,}0468 \quad (4)$$

The best point to take measurements will be the point where the values have higher differences. We can find two different points to take measurements, at high voltages

 ISBN: 978-1-61208-296-7

(1000kHz) and at medium voltages (10kHz). The maximum differences appear at 10kHz.

TABLE VI. INDUCED COIL VOLTAGE AT DIFFERENT FREQUENCIES WHEN THE VOLTAGE IN THE AIR IS FIXED AT 2.8V

Frequency (KHz)	Voltage (V) in Salty water
0,1	2,4
0,5	2,8
1	2,8
10	1,25
25	1,8
50,8	2
100	3,25
145	2,5
248	2,8
500	3,2
750	3,6

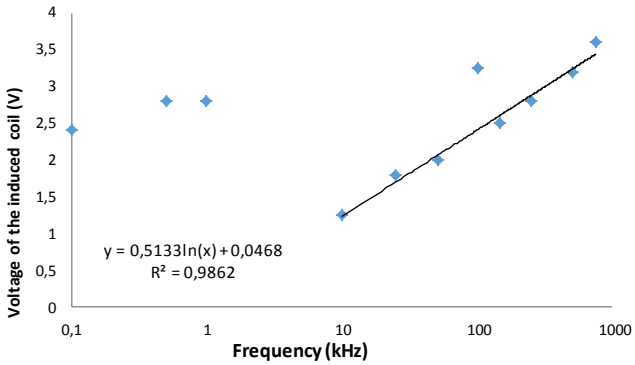


Figure 9. Representation of the data of Table VI.

TABLE VII. COMPARATIVE OF PRICES OF DIFFERENT SENSORS ON THE MARKET.

Name	Fabricant	Physical method	Range on values	Price
WQ-COND	Global Water	Condutive	0 to 200 µS/cm 200 to 2000 µS/cm 2 to 20 mS/cm 20 to 200 mS/cm 200 to 2000 mS/cm	615 €
YSI 5560	YSI	Conductive	-	292 €
PCE-CM 41	PCE Holding GmbH	Conductive	0 to 2000 µS/cm 0 to 20 mS/cm	85,00 €
HI 98309	Hanna Instruments Deutschland GmbH	Conductive	0,000 to 1,999 µS/cm	122,75 €
HI 720122-1	Hanna Instruments Deutschland GmbH	Inductive	0 to 2000 mS/cm	511 €

TABLE VIII. PRICE OF THE COMPONENETS FOR THE SENSOR 1 (SENSOR HALL + SOLENOID)

Component	Prize (€)
Sensor of Hall Effect	1.64
Voltage regulator +5V (1A output current)	14.50
Voltage regulator -5V (1A output current)	2.73
PIC 16f8775.39	5.39
Digital to Analog converter – 8 bits	8.57
Resistors and capacitors	3
Coil solenoid	7

TABLE IX. PRICE OF THE COMPONENETS FOR THE SENSOR 2 (TWO SOLENOIDS

Component	Prize (€)
Voltage regulator +5V (1A output current)	14.50
Voltage regulator -5V (1A output current)	2.73
PIC 16f8775.39	5.39
Digital to Analog converter – 8 bits	8.57
Resistors and capacitors	3
2 x Coil solenoid	14

VI. PRICES COMPARITION

In this section, we are going to make a comparative of the princes of the comertial sensors of conductivity and the developed sensors on this paper. The data of the price of comertail sensors are founded on the different websites of the fabricants, shown in Table VII. Meanwile, the price of the proposed sensors are calculated according to the necessary materials and the electronic components needed to their assembly, shown in Table VIII and Table IX.

We can see that the price can vary quite a lot from one vendor to another. The cheapest costs around 85.00 €, while the most expensive costs around 615 €.

The total price of the fisrt model (Sensor Hall + Solenoid) is 42.83€. By the other hand the total prince of the second model (Two Solenoids) is 48.19.

The new developed sensors are nearly 50% cheapest than the commercial sensors.

VII. CONCLUSIONS AND FUTURE WORK

In this paper, we demonstrated that different environments (air, fresh water and salty water) can produce different alterations in the electromagnetic field, and these alterations can be measured by different methods.

First, by using a solenoid we can measure the values of the electromagnetic field. But the sensibility of this sensor is too low and we need to increase the voltage of the solenoid in order to obtain enough sensibility in the sensor, this produces 2 problems. In the first one, the solenoid heats up and we need to turn off the solenoid between measures because this heat produces interferences (it changes the magnetic field that produces the solenoid). The second one happens because when the magnetic field increases, little changes in the position of the sensor and the solenoid produce erroneous data.

Another way to solve the problem of low sensibility was measuring the electromagnetic field through the induction of voltage in the second coil. In this case we are able to distinguish air from salty water at different frequencies. We have observed that the best frequency to measure is at 10 kHz. In some cases, we have been able to distinguish water with low conductivity from high conductivity.

The advantages of these detection methods versus the traditional ones is that we do not need to put in contact the sensors (the two copper coils or the copper coil with the

 ISBN: 978-1-61208-296-7

magnetic field sensors) with the environment. So, there is no high degradation of the sensor along the time. This is very important because it means that the sensor can be left at any place with no maintenance. Moreover, because the sensor does not have any perishable part or is not consumed during the measurement, the lifespan of the sensor only depends on the energy source.

The main problem in the first set of assays was that there are important changes in the electromagnetic field when the position of the sensor over the solenoid changes. To prevent this in future assays, we propose to create a fixed container for liquids or to make it waterproof and introduce it inside the water. The problem in the second set of assays is that the coils have to be completely isolated. Any hole can make the coils started to drench.

We have several lines to research in future works. Although the highest sensitivity is given at high voltage, this range of measurement has problems. Moreover, the need of an energy source in the environment, where the sensor will be placed, makes us to continue with the assays at low voltages. We will also minimize the size of the coils used to take measurements. In order to achieve this purpose, we will introduce some electronic components that help us to obtain higher values and more sensitivity.

REFERENCES

[1] T. S. Light, S. Licht, A. C. Bevilacqua, and K. R. Morash, The fundamental conductivity and resistivity of water, Electrochemical and Solid-State Letters, vol. 8, issue. 1, 2005. DOI: 10.1149/1.1836121

[2] Z. Jianping, W. Baoliang, H. Zhiyao, and L. Haiqing: A New Electrical Conductance Measurement Instrument Based on Bi-directional Pulsed Voltage Technique. Chinese Journal of Scientific Instrument vol. 26, no. 8, pp. 57–58, 2005

[3] K. M. Rao, A. S. Raghavendra, and K. J. Reddy, Physiology *and molecular biology of stress tolerance in plants*. Kluwer Academic Pub, 2006.

[4] M. Medrano, A. T. Pérez, C. Soria-Hoyo, *Design of a conductivity meter for highly insulating liquids*. Journal of Physics D: Applied Physics, vol. 40, no. 5, pp. 1477-1482, 2007.

[5] Y. Wei, J. Wang, D. Li, Q. Ding, Q, *Design of Intelligent Conductivity Meter Based on MSP430F149*, Third IFIP TC 12 International Conference, CCTA 2009, Beijing, China, October 14-17, 2009, pp. 240-247.

[6] H. Ramos, L. Gurriana, O. Postolache, M. Pereira, P. Girão, *Development and characterization of a conductivity cell for water quality monitoring*, XVIII IMEKO WORLD CONGRESS Metrology for a Sustainaible Development, Sept., 17-22 , 2006, Rio de Janeiro, Brazil.

[7] D. López-Bruna and J. Herrero Isern, *El comportamiento del sensor electromagnético y su calibración frente a la salinidad edáfica*,. Agronomie vol.16, pp. 95-105, 1996.

[8] J.O. Job, M. Rivera, J.L. González, *Algunos usos de la inducción electromagnética en el estudio de los suelos salinos*, Terra Latinoamericana, vol. 16, no. 4 , pp. 309-315, 1998.

[9] Descriptionand characteristics of the conductivity meter CM 35 +:
Avariable at:
www.crisoninstruments.com/file.php?id=79&lang=es

 ISBN: 978-1-61208-296-7