SPACOMM 2014

The Sixth International Conference on Advances in Satellite and Space Communications

ISBN: 978-1-61208-317-9

February 23 - 27, 2014

Nice, France

SPACOMM 2014 Editors

Haibin Liu, China Aerospace Engineering Consultation Center-Beijing, China

SPACOMM 2014

Foreword

The Sixth International Conference on Advances in Satellite and Space Communications (SPACOMM 2014), held between February 23rd-27th, 2014 in Nice, France, continued a series of events attempting to evaluate the state of the art in academia and industry on the satellite, radar, and antennas based communications bringing together scientists and practitioners with challenging issues, achievements, and lessons learnt.

Significant efforts have been allotted to design and deploy global navigation satellite communications systems. Satellite navigation technologies, applications, and services still experience challenges related to signal processing, security, performance, and accuracy. Theories and practices on system-in-package RF design techniques, filters, passive circuits, microwaves, frequency handling, radars, antennas, and radio communications and radio waves propagation have been implemented. Services based on their use are now available, especially those for global positioning and navigation. For example, it is critical to identify the location of targets or the direction of arrival of any signal for civilians or on-purpose applications; smarts antennas and advanced active filters are playing a crucial role. Also progress has been made for transmission strategies; multiantenna systems can be used to increase the transmission speed without need for more bandwidth or power. Special techniques and strategies have been developed and implemented in electronic warfare target location systems.

We take here the opportunity to warmly thank all the members of the SPACOMM 2014 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 SPACOMM 2014. 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 SPACOMM 2014 organizing committee for their help in handling the logistics and for their work to make this professional meeting a success.

We hope that SPACOMM 2014 was a successful international forum for the exchange of ideas and results between academia and industry and for the promotion of progress in the field of satellite and space communications.

We are convinced that the participants found the event useful and communications very open. We also hope the attendees enjoyed the charm of Nice, France.

SPACOMM Advisory Committee:

Stelios Papaharalabos, ISARS/National Observatory of Athens, and Athens Information Technology (AIT), Greece
Piotr Tyczka, Poznan University of Technology, Poland
Michael Sauer, Corning Cable Systems, USA
Ling Pei, Finnish Geodetic Institute, Finland

SPACOMM 2014

Committee

SPACOMM Advisory Committee

Stelios Papaharalabos, ISARS/National Observatory of Athens, and Athens Information Technology (AIT), Greece
Piotr Tyczka, Poznan University of Technology, Poland
Michael Sauer, Corning Cable Systems, USA
Ling Pei, Finnish Geodetic Institute, Finland

SPACOMM 2014 Technical Program Committee

Ashfaq Ahmed, Politecnico di Torino, Italy
Ayman Mahmoud Ahmed, NARSS-Cairo, Egypt
Mohamed Al-Mosawi, University of Portsmouth, UK
Markos P. Anastasopoulos, National Technical University of Athens, Greece
Iva Bacic, University of Zagreb, Croatia
Marco Baldi, Università Politecnica delle Marche - Ancona, Italy
Mark Bentum, University of Twente & ASTRON, The Netherlands
Igor Bisio, University of Genoa - Italy, Italy
Shkelzen Cakaj, Telecom of Kosovo / Prishtina University, Kosovo
Enzo Alberto Candreva, University of Bologna, Italy
Emmanuel Chaput, IRIT-CNRS, France
Bruno Checcucci, Perugia University, Italy
Vittorio Dainelli, Rheinmetall Italia S.p.A. - Rome, Italy
Leonardo Dagui de Oliveira, Escola Politécnica da Universidade de Sao Paulo, Brazil
Francescantonio Della Rosa, Tampere University of Technology, Finland
Felix Flentge, ESA/ESOC HSO-GIB - Darmstadt, Germany
Thierry Gayraud, LAAS-CNRS, Université de Toulouse, France
Mathieu Gineste, Thales Alenia Space, France
Alaa Eldin Saad Hassan, NARSS-Cairo, Egypt
Suk-seung Hwang, Chosun University, Republic of Korea
Konstantinos Kontis, The University of Manchester, UK
Otto Koudelka, TU Graz, Austria
Massimiliano Laddomada, Texas A&M University - Texarkana, USA
Simona Lohan, Department of Electronics and Communications Engineering, Tampere University of Technology, Finland
Krešimir Malaric, University of Zagreb, Croatia
Herwig Mannaert, University of Antwerp, Belgium
Emmanouel T. Michailidis, University of Piraeus, Greece
Marina Mondin, Politecnico di Torino, Italy
Brian Niehöfer, Technische Universität Dortmund, Germany
Nele Noels, University of Gent, Belgium

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

Estimation Quality of High-dimensional Fields in Wireless Sensor Networks

Siyuan Zhou
DET, Politecnico di Torino
Torino, Italy
Email: siyuan.zhou@polito.it

Alessandro Nordio
IEIIT-CNR
Torino, Italy
Email: alessandro.nordio@ieiit.cnr.it

Carla-Fabiana Chiasserini
DET, Politecnico di Torino
Torino, Italy
Email: chiasserini@polito.it

***Abstract*—In the fields of wireless communications, networking and signal processing, systems can be often modeled through a linear relationship involving a random Vandermonde matrix $\mathbf{V}$, and their performance can be characterized through the eigenvalue distribution of the Gram matrix $\mathbf{V}\mathbf{V}^{\mathsf{H}}$. In spite of its key role, little is known about the eigenvalue distribution of such a matrix and only few of its moments are known in closed form. In this work, we obtain a lower and an upper bound to the eigenvalue distribution of $\mathbf{V}\mathbf{V}^{\mathsf{H}}$, as well as an excellent approximation based on entropy maximization. As an application, we consider the case of a wireless sensor network sampling a physical phenomenon to be estimated. We characterize the quality of the estimate through the eigenvalue distribution of $\mathbf{V}\mathbf{V}^{\mathsf{H}}$ by adopting an asymptotic approach, which well suites medium-large scale networks. The proposed method is particularly efficient when dealing with physical phenomena defined over a d-dimensional support, with $d > 2$.**

***Keywords*—sensor networks; Signal estimation; Vandermonde matrices.**

I. Introduction

Recently, random Vandermonde matrices have attracted a great deal of interest since they play an important role in fields such as wireless communications, sensor networks, and image processing. In these contexts, signals estimation systems can often be modeled as [1]–[3]:

$$\mathbf{y} = \mathbf{V}^{\mathsf{H}}\mathbf{a} + \mathbf{n}\,, \tag{1}$$

where the vector $\mathbf{y}$ represents a set of measurements, the vector $\mathbf{a}$ denotes the system input that has to be estimated, $\mathbf{V}$ is a random Vandermonde matrix representing the system transfer function, and $\mathbf{n}$ is the additive noise, uncorrelated with respect to $\mathbf{a}$ and $\mathbf{V}$. Bold lowercase and uppercase letters denote column vectors and matrices, respectively. The conjugate transpose operator is denoted by $(\cdot)^{\mathsf{H}}$, and the identity matrix is denoted by $\mathbf{I}$. In particular, the model in (1) can represent a physical phenomenon sampled by a wireless sensor network composed of nodes randomly deployed on a d-dimensional support. The physical phenomenon is specified by the vector $\mathbf{a}$ which has to be estimated from the set of noisy measurements $\mathbf{y}$, collected by the sensors. In this case, the matrix $\mathbf{V}$ accounts for the positions of the sensor nodes. In [4], it has been shown that the performance of linear estimation techniques can be accurately described through the *eigenvalue distribution* (or *matrix spectrum*) of the Gram matrix associated to $\mathbf{V}$, i.e., $\mathbf{V}\mathbf{V}^{\mathsf{H}}$. In particular, a key role is played by the *asymptotic* spectrum of $\mathbf{V}\mathbf{V}^{\mathsf{H}}$, which is obtained by letting the size of the matrix $\mathbf{V}$ tend to infinity while keeping the ratio of the number of rows to the number of columns constant. Unfortunately, such asymptotic spectrum is still unknown and very few results exist that can shed light on this important issue. For example the results presented in [1], [2] were obtained by using a Monte Carlo approach which, although accurate, turns out to be computationally expensive when the physical phenomenon to be estimated is defined over $d > 2$ dimensions. In this work, we leverage the existing results on the moments of the asymptotic spectrum of $\mathbf{V}\mathbf{V}^{\mathsf{H}}$ and contribute to filling the aforementioned gap by providing (i) a lower and an upper bound to the asymptotic cumulative distribution function of the eigenvalues of $\mathbf{V}\mathbf{V}^{\mathsf{H}}$, and (ii) an approximation of the asymptotic spectrum (both the cumulative and the density functions), which proves to be very accurate. Through such results, we are able to characterize the performance of linear reconstruction techniques by avoiding the need of computationally expensive Monte Carlo methods.

The rest of the paper is organized as follows. First, in Section II we provide some background on the system model and performance metric, as well as on fundamental concepts related to Vandermonde matrices that we use in our analysis. We then present our bounds and approximation of the asymptotic eigenvalue distribution in Section III and Section IV, respectively. Finally, we show numerical results in Section V, and we conclude the paper in Section VI.

II. Signal reconstruction in WSN and random Vandermonde matrices

We consider m sensors sampling a spatially-finite physical phenomenon $s(\mathbf{x})$ (hereinafter also called signal), defined over a d-dimensional hypercube $\mathcal{H} = [0,1)^d$, $d \geq 1$, and with finite energy. The signal can be approximated by its truncated Fourier series expansion so that the sample of the q-th sensor ($q = 1, \ldots, m$) can be modeled as [1]

$$s_q = s(\mathbf{x}_q) = n^{-d/2} \sum_{\boldsymbol{\ell}} a_{\nu(\boldsymbol{\ell})} \mathrm{e}^{\mathrm{j}2\pi\boldsymbol{\ell}^{\mathsf{T}}\mathbf{x}_q} \tag{2}$$

where n is the approximate bandwidth (per dimension) of the field and $\boldsymbol{\ell} = [\ell_1, \ldots, \ell_d]^{\mathsf{T}}$ is a vector of integers, with $\ell_j = 0, \ldots, n-1$, $j = 1, \ldots, d$. The coefficient $n^{-d/2}$ is a normalization factor and the function $\nu(\boldsymbol{\ell}) = \sum_{j=1}^{d} n^{j-1}\ell_j$, maps uniquely the vector $\boldsymbol{\ell}$ into $\{0, \ldots, n^d - 1\}$. The term $a_{\nu(\boldsymbol{\ell})}$ denotes the $\nu(\boldsymbol{\ell})$-th entry of the vector $\mathbf{a} = [a_0, \ldots, a_{n^d-1}]^{\mathsf{T}}$, which represents the approximated signal spectrum, while the vectors $\mathbf{x}_q \in \mathcal{H}$, represents the coordinate of the q-th sampling point (i.e., the position of the q-th sensor), which is assumed to be known. We assume that $\mathbf{x}_q$, $q = 1, \ldots, m$, are i.i.d. random vectors having a generic continuous distribution over the hypercube $\mathcal{H}$. Also, since in general we do not have any information on the signal spectrum $\mathbf{a}$, we assume it has zero

 ISBN: 978-1-61208-317-9

mean and covariance $\sigma_a^2\mathbf{I}$. Without loss of generality and for normalization reasons, we set $\sigma_a^2 = 1$.

The vector of samples $\mathbf{s} = [s_1, \ldots, s_m]^\mathsf{T}$ can be rewritten in a compact form as $\mathbf{s} = \mathbf{V}^\mathsf{H}\mathbf{a}$ where $\mathbf{V}$ is an $n^d \times m$ random Vandermonde matrix whose generic entry

$$(\mathbf{V})_{\nu(\boldsymbol{\ell}),q} = n^{-d/2} \exp\left(-2\pi \mathrm{i} \boldsymbol{\ell}^\mathsf{T}\mathbf{x}_q\right) \tag{3}$$

is randomly distributed on the complex circle of radius $n^{-d/2}$. The vector of noisy samples $\mathbf{y}$ can thus be represented by (1), where $\mathbf{n}$ has zero mean, covariance matrix $\sigma_n^2\mathbf{I}$, and is uncorrelated with $\mathbf{a}$. From the knowledge of $\mathbf{y}$ and of the sampling locations $\mathcal{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_m\}$, the network provides an estimate $\hat{s}(\mathbf{x})$ of the signal $s(\mathbf{x})$. As a performance metric of the signal reconstruction, we take the mean square error (MSE) of the estimate, which can be written in terms of the estimate, $\hat{\mathbf{a}}$, of the signal spectrum $\mathbf{a}$ [1] as

$$\mathcal{M}^{(n,m)} = \mathbb{E}_{\mathbf{a},\mathbf{n},\mathcal{X}} \|\mathbf{a} - \hat{\mathbf{a}}\|^2 / n^d \tag{4}$$

where the average is taken with respect to the subscripted random vectors (see [1] for details). In the literature, many estimators for $\mathbf{a}$ have been proposed. Among these, linear estimators such as the zero-forcing (ZF) and the linear minimum MSE (LMMSE) estimators [4] are commonly employed in signal detection and estimation since their analysis can be often carried out analytically. They are given, respectively, as

$$\begin{aligned} \hat{\mathbf{a}}_{\mathrm{ZF}} &= \left(\mathbf{V}\mathbf{V}^\mathsf{H}\right)^{-1}\mathbf{V}\mathbf{y}, \\ \hat{\mathbf{a}}_{\mathrm{LMMSE}} &= \left(\mathbf{V}\mathbf{V}^\mathsf{H} + \sigma_n^2\mathbf{I}\right)^{-1}\mathbf{V}\mathbf{y}\,. \end{aligned}$$

By using these estimators, the MSE in (4) becomes

$$\begin{aligned} \mathcal{M}_{\mathrm{ZF}}^{(n,m)} &= \frac{\sigma_n^2}{n^d}\mathsf{Tr}\left\{(\mathbf{V}\mathbf{V}^\mathsf{H})^{-1}\right\} \\ \mathcal{M}_{\mathrm{LMMSE}}^{(n,m)} &= \frac{\sigma_n^2}{n^d}\mathsf{Tr}\left\{(\mathbf{V}\mathbf{V}^\mathsf{H} + \sigma_n^2\mathbf{I})^{-1}\right\}. \end{aligned} \tag{5}$$

When the number of sensors, m, and the number of harmonics, n^d, is large with constant ratio $\beta = n^d/m$, the MSE can be tightly approximated by using the asymptotic MSE defined as

$$\mathcal{M}^{\infty} = \lim_{n,m\to\infty} \mathcal{M}^{(n,m)}.$$

For the two filters above, the asymptotic MSE is given by

$$\begin{aligned} \mathcal{M}_{\mathrm{ZF}}^{\infty} &= \sigma_n^2\beta\mathbb{E}_\lambda\left[\lambda^{-1}\right] \\ \mathcal{M}_{\mathrm{LMMSE}}^{\infty} &= \sigma_n^2\beta\mathbb{E}_\lambda\left[(\lambda + \sigma_n^2\beta)^{-1}\right] \end{aligned} \tag{6}$$

where λ is a random variable distributed as the asymptotic spectrum of $\beta\mathbf{V}\mathbf{V}^\mathsf{H}$ (see [1] for details). The knowledge of the distribution of λ (in the following denoted as $f_\lambda(d,\beta,z)$), for the case where the positions $\mathbf{x}_1, \ldots, \mathbf{x}_m$ are i.i.d and uniformly distributed, plays an important role since it allows to compute the asymptotic MSE in (6) also for the case where $\mathbf{x}_1, \ldots, \mathbf{x}_m$ are i.i.d. with generic continuous distribution in $\mathcal{H}$ [5]. Due to the scarcity of results on the behavior of $f_\lambda(d,\beta,z)$, whenever needed, such a distribution has to be evaluated through numerical simulations (i.e., by computing the eigenvalues of several realizations of the $n^d \times n^d$ matrix $\beta\mathbf{V}\mathbf{V}^\mathsf{H}$). Clearly, this is feasible only for fairly small values of n, m and becomes impractical already when $d > 2$.

In order to overcome such a problem, we propose to characterize the distribution by exploiting the only results known in closed-form, i.e, its first few moments. Under the assumption of $\mathbf{x}_1, \ldots, \mathbf{x}_m$ being uniformly distributed over $\mathcal{H}$, we have that the p-th moment of $f_{u,\lambda}(d,\beta,z)$, defined as

$$\mu_p^{(d)} = \int_0^\infty z^p f_{u,\lambda}(d,\beta,z)\,\mathrm{d}z = \sum_{k=1}^{p} \beta^{p-k} \sum_{\boldsymbol{\omega}\in\Omega_{p,k}} v(\boldsymbol{\omega})^d \tag{7}$$

is a polynomial in the variable β [1]. In (7), $\Omega_{p,k}$ is the set of partitions of the set $\mathcal{P} = \{1, 2, \ldots, p\}$ in k subsets, and for any $\boldsymbol{\omega} \in \Omega_{p,k}$, $v(\boldsymbol{\omega}) \in (0, 1]$ is a rational number that can be analytically computed from $\boldsymbol{\omega}$ following the procedure described in [1]. Note that the computational complexity of $\mu_p^{(d)}$, increases with the Bell number of p [1] so only few moments are available. For simplicity, in the rest of the paper we omit the superscript (d) in the expression of the moments.

Using the above expressions, in the following sections we derive a lower and an upper bound to the asymptotic cumulative distribution function (CDF)

$$F_\lambda(d,\beta,z) = \int_0^z f_\lambda(d,\beta,t)\,\mathrm{d}t \tag{8}$$

and a tight approximation of $f_\lambda(d,\beta,z)$ and $F_\lambda(d,\beta,z)$.

III. Bounds to $F_\lambda(d,\beta,z)$

Let Λ be a non-negative random variable with CDF $F_\Lambda(z)$. If $\mathbb{E}[\Lambda] < \infty$, then for any $z > 0$ Markov's inequality states that $\mathbb{P}\{\Lambda \geq z\} \leq \mathbb{E}[\Lambda]/z$, i.e., that

$$F_\Lambda(z) = 1 - \mathbb{P}\{\Lambda \geq z\} \geq 1 - \frac{\mathbb{E}[\Lambda]}{z}\,.$$

Now, let $\Lambda = \lambda^p$ where $\lambda \geq 0$ is distributed as the asymptotic spectrum of $\mathbf{V}\mathbf{V}^\mathsf{H}$, and let us define $\zeta = z^p$. It follows that, for any $p > 0$, we have:

$$\begin{aligned} F_\lambda(d,\beta,z) &= 1 - \mathbb{P}\{\lambda \geq z\} \\ &= 1 - \mathbb{P}\{\lambda^p \geq z^p\} \\ &\geq 1 - \frac{\mathbb{E}[\lambda^p]}{z^p} \\ &= 1 - \frac{\mu_p}{z^p}\,. \end{aligned} \tag{9}$$

If the moments μ_p are available for $p = 1, \ldots, P$, then by (9) a lower bound to $F_\lambda(d,\beta,z)$ can be obtained as:

$$F_\lambda(d,\beta,z) \geq 1 - \min_p \frac{\mu_p}{z^p} \tag{10}$$

An upper bound to $F_\lambda(d,\beta,z)$ can be derived by using the *left-sided Chebychev inequality* [7, chapter 9.1] stating that for any random variable Λ and $\zeta > 0$

$$\mathbb{P}(\Lambda \leq (1-\zeta)\mathbb{E}[\Lambda]) \leq \frac{\mathbb{E}[\Lambda^2] - \mathbb{E}[\Lambda]^2}{\mathbb{E}[\Lambda^2] + (\zeta^2 - 1)\mathbb{E}[\Lambda]^2}\,. \tag{11}$$

Again, by letting $\Lambda = \lambda^p$ in (11), we obtain

$$\mathbb{P}(\lambda^p \leq (1-\zeta)\mu_p) \leq \frac{\mu_{2p} - \mu_p^2}{\mu_{2p} + (\zeta^2 - 1)\mu_p^2}\,.$$

By defining $z^p = (1-\zeta)\mu_p$ (i.e., $\zeta = 1 - z^p/\mu_p$), we can write:

$$\mathbb{P}(\lambda \leq z) = F_\lambda(d,\beta,z) \leq \frac{\mu_{2p} - \mu_{2p}^2}{\mu_{2p} - 2\mu_p z^p + z^{2p}}\,.$$

The above inequality is valid for $\zeta \geq 0$, i.e., $z \leq \mu_p^{1/p}$. When $z > \mu_p^{1/p}$, we assume $F_\lambda(d,\beta,z) \leq 1$. Then, if the moments μ_p are available for $p = 1,\ldots,\lfloor P/2 \rfloor$, we have:

$$F_\lambda(d,\beta,z) \leq \min_p \begin{cases} \frac{\mu_{2p}-\mu_p^2}{\mu_{2p}-2\mu_p z^p + z^{2p}} & \text{if } z \leq \mu_p^{1/p} \\ 1 & \text{else.} \end{cases} \quad (12)$$

IV. Approximation of $f_\lambda(d,\beta,z)$ and $F_\lambda(d,\beta,z)$

The reconstruction of a probability density function from its moments is known as the *Classical Moment Problem*. Unfortunately, the knowledge of a finite set of moments does not guarantee the uniqueness of the solution [8]. In general a good solution must be selected from a solution space according to some cost metric. A method to solve the problem has been proposed in [9], and it is based on the entropy maximization approach.

In practice, an approximation to $f_\lambda(d,\beta,z)$ can be found by maximizing its entropy, under the constraint that the p-th moment of the distribution must be equal to μ_p, for $p = 1,\ldots,P$. More formally, we have to solve the following constrained optimization problem:

$$\begin{aligned} &\max -\int_0^{+\infty} f_\lambda(d,\beta,z)\log f_\lambda(d,\beta,z)\,\mathrm{d}z \\ &\text{s.t.} \\ &\int_0^{+\infty} z^p f_\lambda(d,\beta,z)\,\mathrm{d}z = \mu_p,\ \ p = 0,\ldots,P\,. \end{aligned} \quad (13)$$

When the number of known moments is low, i.e., for $P \leq 2$, $f_\lambda(d,\beta,z)$ can be reconstructed analytically [11]. If only μ_1 is known, we have $f_\lambda(d,\beta,z) = \exp(-z)$. If μ_1 and μ_2 are known, $f_\lambda(d,\beta,z)$ behaves as Gaussian function and is given by $f_\lambda(d,\beta,z) = \exp(-(1+a+bz+cz^2))$, where a, b and c are solutions of the following multivariate equations:

$$\begin{aligned} &\mathrm{e}^{-1-a}\left(\tfrac{\frac{b^2}{2}+c}{4}\sqrt{\tfrac{\pi}{c^5}}\mathrm{e}^{\frac{b^2}{4c}}\left(1-\mathrm{erf}\left(\tfrac{b}{2\sqrt{c}}\right)\right) - \tfrac{b}{4c^2}\right) = \mu_2 \\ &\mathrm{e}^{-1-a}\left(\tfrac{1}{2c} - \tfrac{b}{4c}\sqrt{\tfrac{\pi}{c}}\mathrm{e}^{\frac{b^2}{4c}}\left(1-\mathrm{erf}\left(\tfrac{b}{2\sqrt{c}}\right)\right)\right) = \mu_1 \\ &\mathrm{e}^{-1-a}\sqrt{\tfrac{\pi}{4c}}\mathrm{e}^{\frac{b^2}{4c}}\left(1-\mathrm{erf}\left(\tfrac{b}{2\sqrt{c}}\right)\right) = 1 \end{aligned}$$

However, for a larger number of moments, i.e., $P > 2$, (13) is in general intractable analytically. We therefore solve it numerically by resorting to the stable algorithm in [8], which is based on the discretization of the integrals through an N-points Gaussian quadrature rule [10]. The method consists in approximating the function to be integrated with the product of a polynomial function and a weighting function $w(x)$, and then in discretizing the latter two. Specifically, we can write:

$$-\int_0^{+\infty} f_\lambda(d,\beta,z)\log f_\lambda(d,\beta,z)\,\mathrm{d}z \approx -\sum_{j=1}^{N} w_j f_j \log f_j$$

and

$$\int_0^{+\infty} z^p f_\lambda(d,\beta,z)\,\mathrm{d}z \approx \sum_{j=1}^{N} w_j z_j^p f_j,\ \ p = 0,\ldots,P$$

where, for $j = 1,\ldots,N$, $f_j = f_\lambda(d,\beta,z_j)$ while z_j and w_j are, respectively, the abscissae of the Gaussian quadrature rule and the corresponding values of the weighting function.

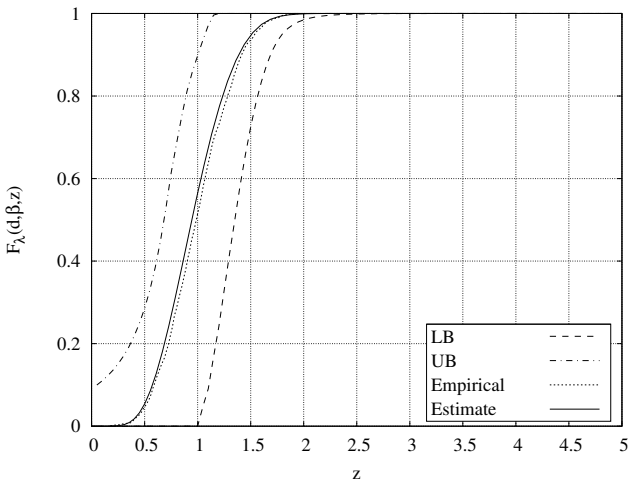


Fig. 1. Comparing bounds, maximum entropy approximation of the CDF, and the empirical distribution, with $d = 1$ and for $\beta = 0.1$.

V. Estimation Performance

We validate our proposed approach by comparing the bounds and approximation against the empirical distribution. The latter has been obtained by computing the eigenvalues of several realizations of $\mathbf{V}\mathbf{V}^{\mathsf{H}}$ for $n = 100$ and a number of sensors $m = n/\beta$, while the maximum entropy approximation has been computed by using the first 12 moments of the asymptotic eigenvalue distribution. All results refer to the case where the phases of the Vandermonde matrix are uniformly distributed over $[0,1)^d$; note, however, that other phase distributions could be considered as well by leveraging the results in [1], [2].

The curves depicted in Figure 1 depict the CDF $F_\lambda(d,\beta,z)$, and have been obtained with $d = 1$ and $\beta = 0.1$. Observe that our bounds follow the behavior of the empirical distribution very well, and there is an excellent match between the latter and the maximum entropy approximation.

Figure 2 compares our approximation to the probability density function, $f_\lambda(d,\beta,z)$, against the empirical results, when $d = 1$ and for $\beta = 0.2$ (top plot) and $\beta = 0.7$ (bottom plot). As expected, in this case the differences between the approximation and the empirical results are more evident than in the case of the CDF, however our approximation still shows to be very tight, even for β as high as 0.7. In addition, the top plot compares the analytical solution of (13), obtained using μ_1 and $\mu_{1,2}$, to our approximation and the empirical results. Clearly, the higher the number of considered moments, the better the approximation accuracy with respect to the empirical results. We also stress that, by definition of β, meaningful values of such a parameter are limited to the $[0,1]$ interval, as the number of sensors (m) should always outnumber the signal harmonics (n^d).

Figure 3 shows the asymptotic MSE, $\mathcal{M}^\infty$, achieved by the ZF and LMMSE filters for $d = 1, 4$ and for a noise variance $\sigma_n^2 = 0.01$. The curves were obtained by computing (6) where the distribution of λ was approximated by solving the problem in (13). As expected, the LMMSE filter performs better than the ZF filter since it minimizes the MSE. Also, the figure shows that given number of harmonics per dimension, n,

 ISBN: 978-1-61208-317-9

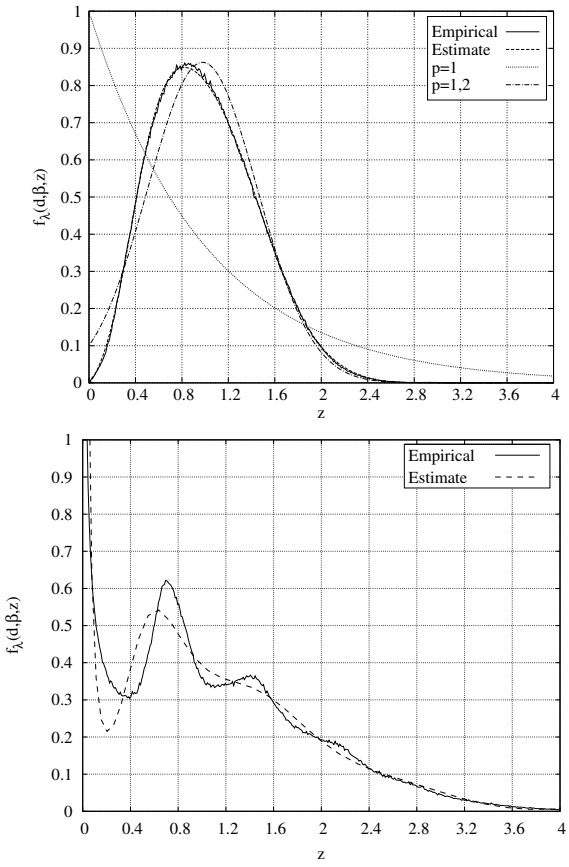


Fig. 2. Comparison between the empirical results and the maximum entropy approximation of the probability density function, with $d = 1$ and for $\beta = 0.2$ (top) and $\beta = 0.7$ (bottom). In the upper plot, it is shown also the analytical solution with p=1, and p=1,2.

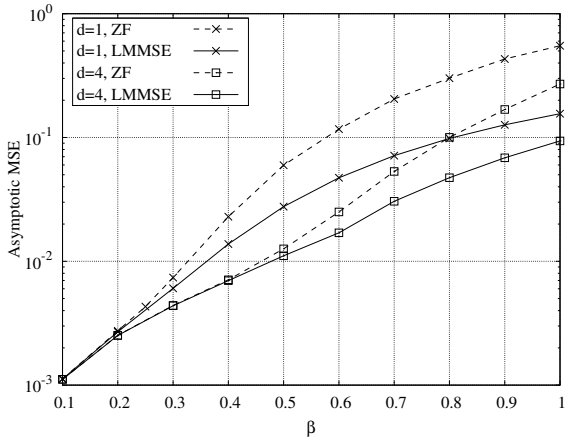


Fig. 3. Comparison between the asymptotic MSE, $\mathcal{M}^\infty$, achieved by the ZF and the LMMSE filters, for $d = 1, 4$ and $\sigma_n^2 = 0.01$.

the asymptotic MSE decreases with the number of available measurements m, i.e., it increases with $\beta = n^d/m$. More importantly, we remark that the maximum entropy approach proposed here, allows to efficiently estimate $f_\lambda(d, \beta, z)$ and $\mathcal{M}^\infty$ for any d, thus avoiding to resort to the numerical computation of the eigenvalues of large matrices.

VI. Conclusions and Future Work

We studied the asymptotic eigenvalue distribution of the Gramian of random Vandermonde matrices, which has an important role in determining the performance of many systems for signal estimation. In particular, we derived a lower and an upper bound to the asymptotic cumulative distribution function. Additionally, we provided an approximation of both the cumulative distribution and the probability density functions, which showed to be very accurate, without applying the cumbersome computation of empirical results.

Our future work will mainly focus on the application of this approximation method to the achievable mutual information of systems, when the channel behavior can be represented by Vandermonde matrix [2]. In this case, we are able to compute the achievable mutual information by deriving the eigenvalue distribution through its moments. Other possible extensions will consider the mutual information in multi-user MIMO systems, and multifold scattering scenarios.

Acknowledgment

This paper was made possible by NPRP grant $\sharp 5 - 782 - 2 - 322$ from the Qatar National Research Fund (a member of Qatar Foundation). The statements made herein are solely the responsibility of the authors.

References

[1] A. Nordio, C.-F. Chiasserini, and E. Viterbo, "Reconstruction of multidimensional signals from irregular noisy samples," *IEEE Trans. on Signal Processing,* Vol. 56, No. 9, Sept. 2008, pp. 4274–4285.

[2] O. Ryan, and M. Debbah, "Asymptotic Behavior of Random Vandermonde Matrices with Entries on the Unit Circle", *IEEE Trans. on Information Theory*, Vol. 55, No. 7, July 2009, pp. 3115–3147.

[3] A. Nordio, G. Alfano, C.-F. Chiasserini, and A. M. Tulino, "Asymptotics of Multifold Vandermonde Matrices with Random Entries,", *IEEE Trans. on Signal Processing,* Vol. 59, No. 6, June 2011, pp. 2760–2772.

[4] A. Nordio, C.-F. Chiasserini, and E. Viterbo, "Performance of linear field reconstruction techniques with noise and uncertain sensor locations," *IEEE Trans. on Signal Processing,* Vol. 56, No. 8, Aug. 2008, pp. 3535–3547.

[5] A. Nordio, and C.-F. Chiasserini, "Field Reconstruction in Sensor Networks with Coverage Holes and Packet Losses," *IEEE Trans. on Signal Processing,* Vol. 59, No. 8, Aug. 2011, pp. 3943–3953.

[6] G. H. Tucci, and P. A. Whiting, "Eigenvalue Results for Large Scale Random Vandermonde Matrices with Unit Complex Entries," *IEEE Trans. on Information Theory*, Vol. 57, No. 6, June 2011, pp. 3938–3954.

[7] R. Tempo, G. Calafiore, and F. Dabbene, *Randomized Algorithms for Analysis and Control of Uncertain Systems*, Springer-Verlag, 2005.

[8] K. Bandyopadhyay, K. Bhattacharya, P. Biswas, and D. A. Drabold, "Maximum Entropy and the Problem of Moments: A Stable Algorithm," *Physical Review E 71, 057701*, 2005.

[9] L. R. Mead, and N. Papanicolaou, "Maximum Entropy in the Problem of Moments," *J. Math. Phys. 25, 2404*, Aug. 1984, pp. 2404–2417.

[10] M. Hazewinkel, "Gauss Quadrature Formula," *Encyclopedia of Mathematics*, Springer, 2001.

[11] F. Socheleau, C. Laot, and J. Passerieux, "Concise Derivation of Scattering Function from Channel Entropy Maximization," *IEEE Trans. on Communications*, Vol. 58, No. 11, Nov. 2010, pp. 3098–3103.

 ISBN: 978-1-61208-317-9

Estimation of Vertical Ionosphere From The Oblique Sounding Measurement Using HF Radar System

Jin-Ho Jo, Moon-Hee You, Yong-Min Lee, Cheol-Oh Jeong
Satellite & Wireless Convergence Research Department,
Electronics and Telecommunications Research Institute (ETRI),
Daejeon City, Rep. of Korea
E-mail: jhjo@etri.re.kr, moon@etri.re.kr, ymlee01@etri.re.kr, cojeong@etri.re.kr

Abstract— **Many of radio wave applications require real-time determination of the 2-dimensional structure of the ionospheric electron density along the path, or at least of the mid-path electron density profile. Since it is often impractical to deploy a vertical ionosonde at the point of interest, such as across oceans, the idea of reconstructing an average mid-path profile from oblique incidence measurements is very appealing. In this paper, we present methods for estimate vertical ionosphere profile without vertical ionospheric measurement. This can be achieved by ionogram conversion from the oblique to equivalent vertical at the midpoint between the measurement stations. To get vertical ionogram from the measured oblique ionogram, two steps processing are necessary. The first step is a trace extraction from the ionogram and vector tracking algorithm is used in this phase. Second step is an ionogram conversion from the oblique to equivalent vertical. In order to verify ionogram conversion algorithm, converted vertical ionograms are compared with the measured ionograms. The comparison results show that ionogram conversion methods can be used for the midpoint vertical ionogram estimation without vertical ionospheric measurements.**

Keywords-Ionosphere; Ionogram; Ionosonde; Vertical sounding; Oblique sounding; Trace extraction; Ionobgram conversion; Vector tracking.

I. INTRODUCTION

The ionosphere can be the largest source of error in GPS positioning and navigation. Furthermore, under certain condition, amplitude fading and phase scintillation effects can cause loss of carrier lock and intermittent GPS receiver operation. The ionosphere effects radio propagation especially HF radio communication. The presence of the free electrons in the ionosphere effects radio signals across several bands of radio frequency.

For the ionospheric research, Korean Space Weather Center (KSWC) installed ionosphere measurement radar system, called ionosonde [1], in the place of Jeju and Icheon. Both stations measure vertical ionospheric condition of each sky respectively. The vertical ionosphere measurement can provide the important ionospheric parameters such as critical frequency, virtual height and electron density, for ionoshperic research. Meanwhile oblique ionosphere measurement has the ability to detect the ionosphere over sea and other terrain where it is not practical to deploy vertical ionosonde and provide more information with less transmitting and receiving devices. In a network for ionospheric monitoring, N vertical incident ionosondes can provide the ionospheric information of N locations overhead, but N transmitters and M receives of the oblique incident ionosondes, which are separated from each other, have the potential of characterizing the ionosphere in the vicinity of N×M reflecting points. Therefore, the oblique incident ionosondes still have kept developing by now [2].

This paper describes methods for estimate vertical ionospheric condition without ionosphere measurement. This can be achieved by ionogram conversion from oblique measurement to equivalent vertical at the midpoint between the stations [5], [6]. The conversion algorithm presented in this paper provides not only vertical profile of F layer but also provide information for sporadic E layer occurrence of ionosphere.

In this paper, ionosphere is reviewed first at Section 2. The ionosphere measurement radar system called ionosonde is reviewed at Section 3. Ionosphere measurement status in Korea is reviewed at Section 4. Method for estimate vertical ionospheric condition at midpoint from the oblique sounding measurements at both stations is presented at Section 5. Validation results for estimated vertical ionospheric condition at midpoint is reviewed Section 6. Finally research activities are summarized in Section 7.

II. IONOSPHERE

The ionosphere is a region of the upper atmosphere, from about 85km to 600km altitude. It plays an important part in atmospheric electricity and forms the inner edge of the magnetosphere. It has practical importance because, among other functions, it influences radio propagation to distant places on the Earth.

The ionosphere effects radio propagation from the extremely low frequencies to super high frequencies. Below 30MHz the ionosphere is an essential part of the propagation, whereas above 30MHz the ionosphere is a source of band pollution particularly at night in the LF and MF bands (30kHz to 3MHz) and system disruption for Earth-space communications, such as navigation systems [3].

The D layer is the innermost layer, 60 km to 90 km above the surface of the Earth. The E layer is the middle layer, 90 km to 120 km above the surface of the Earth. The F layer or

 ISBN: 978-1-61208-317-9

region, also known as the Appleton layer, extends from about 200 km to more than 500 km above the surface of Earth. The Figure 1 shows the structure of ionosphere. At night, the F layer is the only layer of significant ionization present, while the ionization in the E and D layers is extremely low. During the day, the D and E layers become much more heavily ionized, as does the F layer, which develops an additional, weaker region of ionization known as the F1 layer. The F2 layer persists by day and night and is the region mainly responsible for the refraction of radio waves.

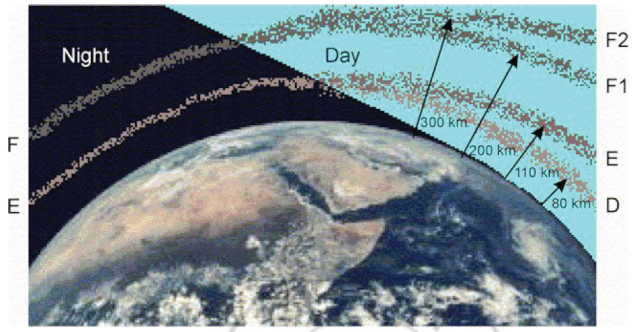


Figure 1. Struectre of ionosphere at day and night

III. IONOSONDES

An ionosonde is a special radar system for the examination of the ionosphere. The changing state of the ionosphere is generally monitored by networks of vertical ionosondes. The transmitter of ionosonde sweeps all or part of the HF frequency range, transmitting short pulses. These pulses are reflected at various layers of the ionosphere, at heights of 100-400 km, and their echo are received by the receiver and analysed by the control system. The result is displayed in the form of an ionogram, a graph of reflection height (actually time between transmission and reception of pulse) versus carrier frequency. An ionosonde consists of;

- A high frequency (HF) transmitter, automatically tuneable over a wide range. Typically the frequency coverage is 0.5-23MHz or 1-40MHz, though normally sweeps are confined to approximately 1.6-12MHz.
- A tracking HF receiver, which can automatically track the frequency of the transmitter.
- An antenna with a suitable radiation pattern, which transmits well vertically upwards and is efficient over the whole frequency range used.
- Digital control and data analysis circuits.

In Korea, model DPS-4D ionosonde are used for ionosphere layer measurements. Figure 2 shows the DPS-4D ionosonde and receive antenna installed in Korea. Table-1 shows the characteristics of DPS-4D ionosonde [1]. The DPS-4D is the latest digital ionosonde that the University of Massachusetts Lowell Centres for Atmospheric Research developed during 2004-2008. The DPS-4D ionosonde uses one simple crossed delta or rhombic antenna for transmission, and an array of four small crossed loops for reception. The DPS-4D system compensates for a low power transmitter (300 W vs. 10 kW for previous systems) by employing intra-pulse phase coding, digital pulse compression and Doppler integration.

Figure 2. DPS-4 ionosonde and receive antenna

TABLE I. MAIN CHARACTERISTICS OF DPS-4D

Parameters	Characteristics
Frequency scan	0.5 ~ 30MHz
Frequency synthesis	Fully digital
Pulse width	533us (16 chips of 33us)
Peak pulse power	2 channels @150W each
Doppler range	±3Hz ~ ±50Hz
Doppler resolution	0.0125 ~ 12.5Hz
TX antenna type	Turnstile Delta or Rhombic
RX antenna type	Active crossed loop

IV. IONOSPHERE MEASUREMENTS IN KOREA

In the past, the vertical incidence ionosonde has been the mainstream for ionospheric detection, due to its ability to accurate measure the echo group range as a function of frequency and to estimate the ionospheric critical frequency, virtual height, and electron density profile so on. However, oblique incidence ionosonde has not been widely applied. The oblique incidence ionosonde can monitor the ionosphere over ocean, marsh, desert and so on, where it is not practical to deploy vertical incidence sounder. Oblique sounding may be used to obtain information about the ionosphere near the mid-point between two ionosonde sites. This makes it useful for recovering ionospheric data at a point where vertical sounding cannot be conducted.

There are two ionosonde stations in Korea, named Jeju station located in 33.4°N, 126.3°E and Icheon station located in 37.1°N, 127.5°E. The ionosondes are operation in the frequency sweep mode, transmitting the phase coded pulse train by the transmit antenna. The GPS receivers are applied in the ionosondes for the time and frequency synchronization between the transmitter and receiver. The output of ionosonde measurements are ionograms. An ionogram is a graph of the virtual height of the ionosphere plotted against frequency. Ionogram is often converted into electron density

 ISBN: 978-1-61208-317-9

profiles. Data from ionograms may be used to measure changes in the Earth's ionosphere due to space weather events.

In the ionosonde stations Jeju and Icheon, oblique ionograms 1 to 16MHz operating frequencies with a 25 kHz frequency step are recorded every 30 minutes each. The height bin of vertical ionogram is from 80 to 1280km with 2.5km step. The Figure 3 shows the oblique sounding configuration of two ionosonde stations to get midpoint ionogram information.

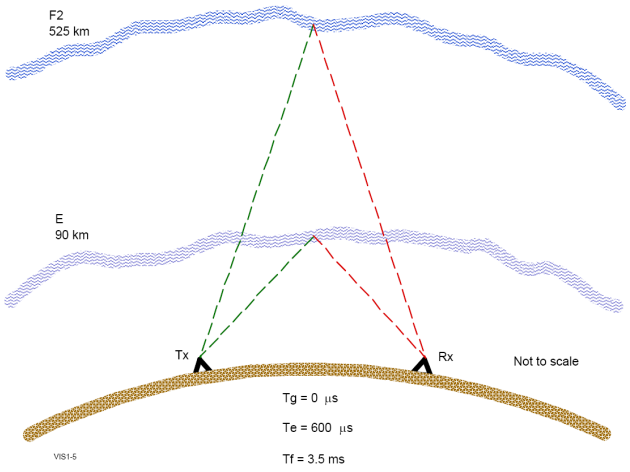


Figure 3. Oblique sounding configuration for the midpoint ionogram measurement

The ground range between Jeju and Icheon stations have is about 430km. The one hop ionospheric reflecting point of oblique propagating path from Jeju to Icheon stations is near the Gochang in Korea. If ionosonde in Jeju station receive pulse signal transmitted from the ionosonde in Icheon station, the strongest receiving signal is a one hop reflected signal at the ionosphere near the Gochang. So oblique sounding ionogram measured at receiving station has information of ionospheric condition at the midpoint between the stations.

The Figure 4 is a map showing the location of Digisonde station Jeju and Icheon and one hop ionospheric reflecting point of oblique propagating path between two stations (midpoint).

Figure 5 shows the oblique ionogram measured at Jeju station. The x axis of ionogram is a frequency with MHz scale and y axis of ionogram is a height with km scale. The color of ionogram shows a direction of signal arrival. When Jeju station receives signal from the Icheon station, arrival azimuth angle is 330 degree and this direction is displayed as a sky blue color in the ionogram. The sweep frequency of ionosonde is from 1 to 16MHz and measurement hight range is from 8 to 1280km. There are two sky blue color signal clusters in the lower part of figure. The left one of the two signal clusters is an Ordinary signal path and right one is an extraordinary signal path. Extraordinary signal is same as Ordinary except polarization was rotated by the geomagnetic field during the ionospheric propagation. The upper signal cluster is a double hop signals.

Figure 4. Location of ionosinde stations and midpoint in Korea

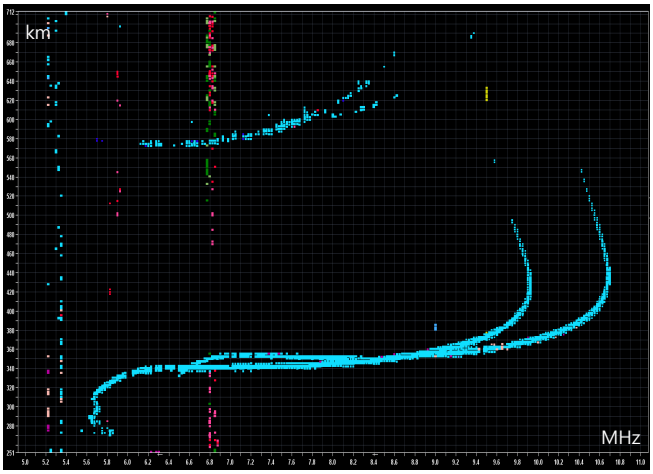


Figure 5. Oblique ionogram measured at Jeju

V. VERTICAL IONOGRAM AT MIDPOINT

As mentioned in section 4, vertical ionogram is more useful than oblique ionogram for the ionospheric research due to its ability to accurate measure the echo group range as a function of frequency and to estimate the ionospheric critical frequency, virtual height, and electron density profile, so on. In this section the idea getting vertical ionogram of the

 ISBN: 978-1-61208-317-9

midpoint from the measured oblique ionogram at two ionosonde stations are described. To get vertical ionogram from the measured oblique ionogram, two processing steps are necessary. First step processing is trace extraction from the ionogram and second step processing is ionogram conversion from oblique to vertical.

A. Trace extraction from the ionogram

Before ionogram conversion from oblique to vertical, it needs to extract trace from the oblique ionogram. The trace extraction reduces data size of oblique ionogram and make easy to handle conversion calculation. The problem of reliable trace extraction is a difficult one. Typically only vertical incidence ionograms have been collected regularly because techniques for trace extraction from such have been developed. The variety of trace shapes in oblique sounding due to different ionospheric conditions, bease-lines and ionospheric gradients, adds an order of difficulty to the problem and there seems to be as yet little work published in the open literature on the problem [4].

We developed new algorithm called 'vector tracking' for the trace extraction from the oblique ionogram data. This algorithm can extract ionogram trace well even in week echo signal and noisy environments. The starting point of trace extraction is maximum amplitude point in scattered data of the oblique ionogram. In vector tracking algorithm, amplitude value of eight directions from the maximum amplitude point are compared and maximum amplitude direction is selected as a next trace point. The eight directions are angles of 0, 45, 90, 135, 180, 225, 270, and 315 degree from the selected trace points. Figure 6 shows a trace extraction method in vector tracking algorithm. Four amplitude value of echo signal in eight directions in the ionogram are added and compared each other and direction of maximum amplitude value is selected as a next trace point.

Figure 7 shows example of trace extraction from the ionogram using 'vector tracking' algorithm. Red line in the figure is an extracted trace of ionogram from the starting point in right hand. Blue line in the figure is an extracted trace of ionogram from the starting point in left hand.

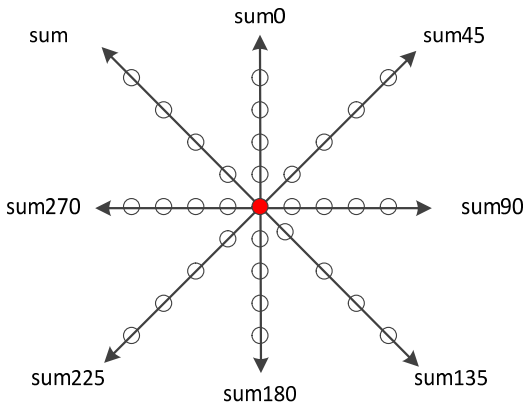


Figure 6. Trace extraction method in vector traking algorithm

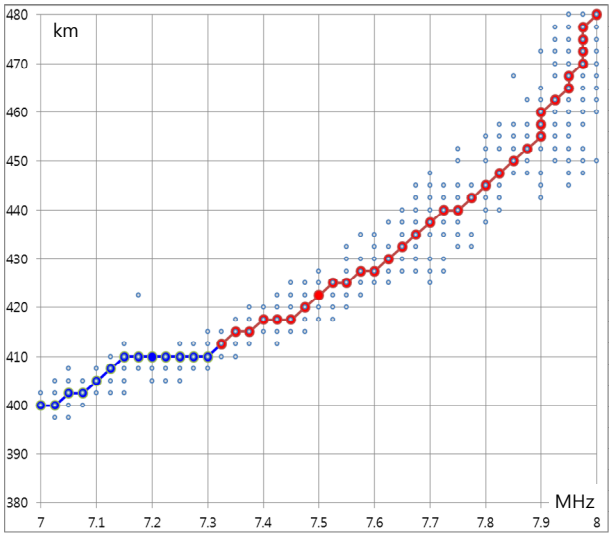
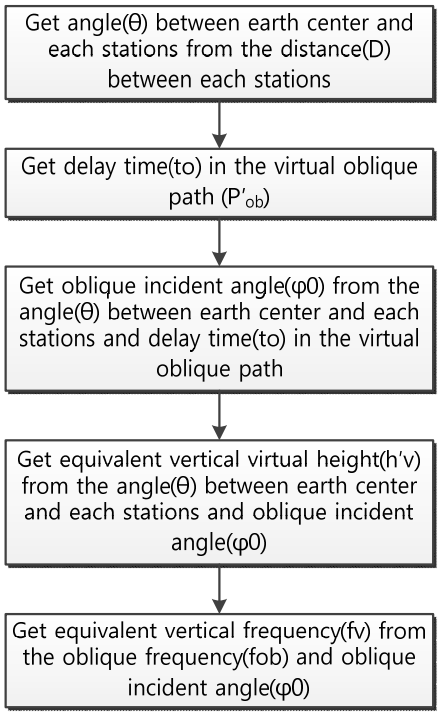


Figure 7. Ionogram trace extraction by vector tracking algorithm

B. Ionogeam conversion from oblique to vertical

The distance between Jeju and Icheon stations is less than 500km. Therefore, it is acceptable to assume ionosphere layers are flat and there are no electron confliction and no geomagnetic disturbance between the stations. In this case, conversion algorithm can be expressed by secant law theory of Berit and Tuve [5] and theory of Martyn's equivalent path [6]. The vertical frequency (f_v) and virtual height (h'_v) can be derived from the oblique frequency (f_{ob}) and virtual path (P'_{ob}). The detailed conversion algorithms can be described in following steps.



 ISBN: 978-1-61208-317-9

From the above steps, we can get equivalent vertical frequency (f_v) and vertical virtual height (h'_v) from the oblique frequency (f_{ob}) and the virtual oblique path (P'_{ob}) between the stations. The equations are as following.

$$f_v = \frac{f_{ob}}{1.002}\cos\left(\sin^{-1}\left(\frac{428.8748}{P'_{ob}}\right)\right) \quad (1)$$

$$h'_v = -3.6097 + \frac{214.4374}{\tan\left(\sin^{-1}\left(\frac{428.8748}{P'_{ob}}\right)\right)} \quad (2)$$

The Figure 8 shows a measured oblique ionogram at both station and converted equivalent vertical ionogram of the midpoint according to the (1) and (2) in above.

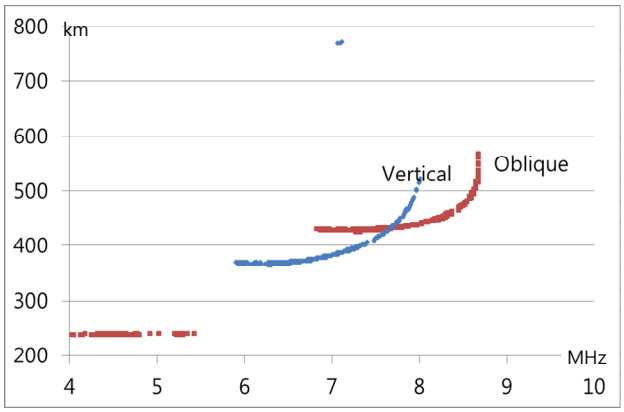


Figure 8. Measured oblique ionogram and converted vertical ionogram

VI. Validation of Algorithm

It needs to validate equivalent vertical ionogram of midpoint between two stations converted from the measured oblique ionogram. The best way is compare converted equivalent vertical ionogram with a measured vertical ionogram at the midpoint. But this method is not practical because it needs new ionosonde installation at the midpoint. The alternative way is compare measured vertical ionogram at both stations. This method seems to be reasonable because distance between the stations is just 430km and measurement interval of each station is short as 8 minutes. So if ionosphere condition is calm, alternative method can be a good choice for the validation.

Figure 9 shows a measured vertical ionograms at Jeju, Icheon stations and equivalent vertical ionogram of midpoint. Equivalent vertical ionogram in the figure is located between the two station ionograms and shape is similar with both ones. This means that ionogram conversion equations (1) and (2) in previous chapter are validated and this conversion can be used for the vertical ionogram estimation without measurement at the midpoint.

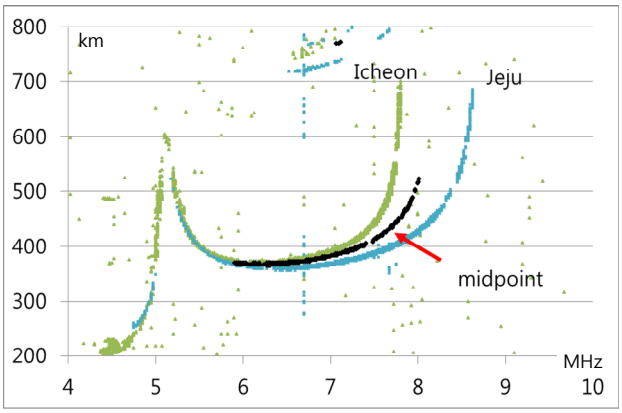


Figure 9. Ionograms of Jeju, Icheon and midpoint

VII. Conclusion and Futher Works

In this paper, we present methods for estimate ionospheric condition without ionosphere measurement. This can be achieved by ionogram conversion from the oblique to equivalent vertical at the midpoint between the measurement stations. First step is a trace extraction from the measured oblique ionogram of the receiving station. The vector tracking algorithm was used in this phase. Second step is an ionogram trace conversion from the oblique to vertical. The equivalent vertical conversion algorithm was used in this phase. In order to verify the conversion algorithm, converted vertical ionogram of midpoint was compared to the measured vertical ionograms of both receiving stations. The result shows that conversion algorithm can be used for the vertical ionogram estimation at the midpoint.

In Korea, we have plans to install new ionosondes in the most east and west points for the full mesh ionosphere measurement network. Then we can get not only 4 stations vertical ionograms, but also midpoints vertical ionogram from the 4 stations measurements using these algorithms. Also we will extend this study to over country sea area. We have plans to cooperate with Japan and China for this research.

References

[1] B. W. Reinisch, Digisonde 4D Technical Manual (Version 1.0), Lowell Digisonde International, 2009.

[2] Chen Gang, “Application of the oblique ionogram as vertical ionogram,” Science China Technical Sciences, vol. 55, no. 5, pp. 1240-1244, 2012.

[3] R. Bamford, The Oblique Ionospheric Sounder (Project Final Report), Rutherford Appleton Laboratory, 2000.

[4] Matthew Roughan, “Trace Extraction From Oblique Ionograms,” Applied MAchine Vision, ’96, Cincinnati, Ohio, USA, 1996.

[5] G. Breit and M. A. Tuve, “A Test of the Existence of the Conducting Layer,” Physical Review, vol. 28, no. 3, pp. 554-575, 1926.

[6] D. P. Martyn, “The Propagation of Medium Radio Waves in the Ionosphere,” Proceedings of the Physical Society, vol. 47, no. 2, pp. 323-339, 1935.

 ISBN: 978-1-61208-317-9

Ultra Low Power CMOS Phase Locked Loop synthesizer for Very High Frequencies

Nayera Ahmed
Communication and Electronics Department.
Alexandria University, Faculty of Engineering
Alexandria, Egypt

Akram Malak
Communication and Electronics Department.
Ain-Shams University, Faculty of Engineering
Cairo, Egypt

Abstract—**This paper describes the design of an essential component in wireless transceivers, the frequency synthesizer. The synthesizer is implemented using Phase Locked Loop (PLL). Second order PLL, type II, with a reference frequency 10MHz is designed using 180nm analog CMOS process technology. The synthesizer generates signals in frequency range of 10-100 MHz. The simulated power consumption of the system is 37µW with a deviation from the true periodicity; root mean square periodic jitter is in range of 5 ps.**

Keywords- PLL; CMOS; ,frequency; synthesizer.

I. INTRODUCTION

The main job of frequency synthesizer is to create a set of frequencies multiple of a reference frequency. Such specification can be made using three techniques: the look up table synthesizer, the direct/indirect synthesizer and Phase-Locked Loop (PLL) synthesizers [1]. PLL frequency synthesizers offer high level of stability and accuracy determined by the crystal oscillator; they are also easy to control using a digital control circuit. PLL plays a major role in the field of communication systems [2][3], where the major target is saving battery energy, PLL is the most power consumer block in the transceiver [4]. Radio Frequency, nowadays, is used as the intermediate between the sensor unit and the computer for biomedical applications, mostly in the muscular stimulation (heart diseases) [5]. Any circuit within the chip should consume ultra-low power and area. In this paper, a PLL is proposed as a frequency synthesizer, based on Phase/Frequency Detector (PFD), Charge Pump (CP), Voltage Control Oscillator (VCO) and Frequency Divider (FD), which are implemented using 0.18µm CMOS. In order to minimize the power consumption, a lower supply voltage is used. N-divider is used to synthesize the Reference Frequency (F_{ref}). Targeting high output frequency needs high power consumption.

In this paper, we discuss a compromise VHF PLL with a limited current budget. The paper is organized as follows. Section II presents the theoretical background. Circuit design and simulation results are discussed in Section III. In Section IV, the summary of system performance is presented. Conclusion and future work are discussed in Section V.

II. SYSTEM DESIGN AND BEHAVIORAL SIMULATION

In this section, a behavioral model is introduced to define the system specs and ensure the system stability. PFD detects the phase and frequency error. CP and Loop Filter compute the phase error and convert it to average voltage information used by VCO, which creates a frequency synchronized with $F_{ref.}$.

The system is modeled in Figure 1 using linear approaches; PFD/CP detects the error in the phase. VCO generates the output frequency from the input controlled voltage.

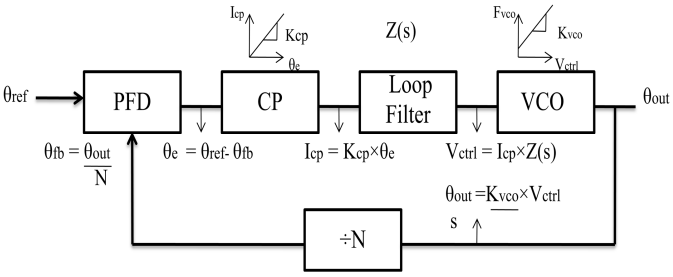


Figure 1. Model of PLL

From the theory of the Feedback [5]:

$$H_{CL} = \frac{\theta_{out}}{\theta_{ref}} = \frac{N \times H_{OL}}{1 + H_{OL}} \quad (1)$$

where H_{CL} is the closed loop gain, while H_{OL} is the open loop gain.

$$H_{OL} = \frac{K_{CP} \times Z(s) \times K_{VCO}}{s \times N} = \frac{K_{vco}\, I_p\, R_p}{2\pi N} \times \frac{s + \frac{1}{R_p C_p}}{s^2} \quad (2)$$

where;
K_{CP}: The charge pump gain (A/rad).
K_{VCO}: The VCO gain gain (Hz/V).
Z(s): The Loop Filter (LF) impedance

To ensure the system stability, from (2), a phase of H_{OL}< 180° should be guaranteed at the unity gain frequency (f_t). PLL second order has two poles at DC (phase $H_{OL}(0)$ = 180°); a zero is introduced at low frequency to maintain the system stability. In contrast, Phase Margin (PM)>0, i.e., the system stability is guaranteed as shown in Figures 2. Figure 3 shows the weak overshoot of the step response.

 ISBN: 978-1-61208-317-9

The maximum overshoot is often used as a measure of the relative stability of the control system. The settling time is

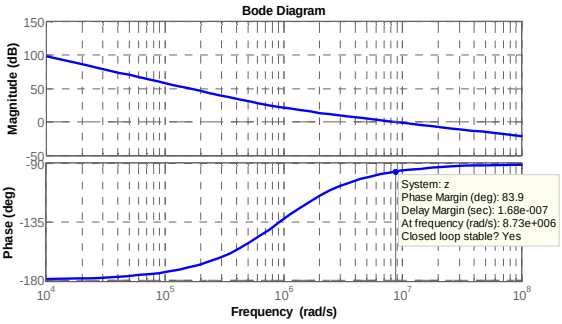


Figure 2. Phase Margin

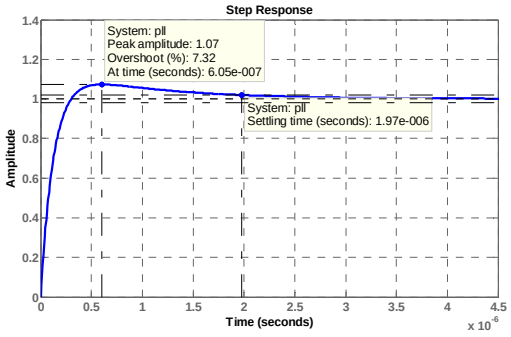


Figure 3. Step response

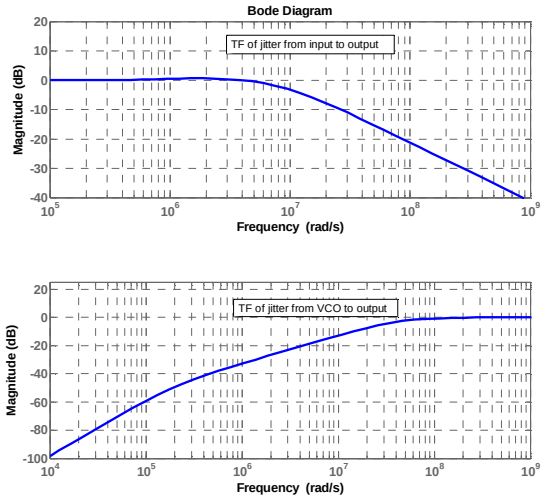


Figure 4. Transfer of jitter

$2\mu s \propto \frac{1}{5f_t}$. In our application, the time needed to achieve lock condition (constant input output phase difference), is not critical; thus, the unity gain frequency is used to be in order of 1MHz.The transfer function for Type II PLL has a low pass characteristic; as a result, slow jitter at the input propagates to the output, but fast jitter does not. On the other hand, the phase noise VCO transfer function is a high pass filter introduced in (3). Figure 4 presents the phase noise transfer functions of the input reference and the VCO. For that reason, in our design, the CP consumption has been reduced, while increasing the VCO current for a better phase noise at the frequency of interest, while respecting the estimated low power consumption budget.

$$\frac{\varphi_{FBDIV}}{\varphi_{ref}}(s) = \frac{s^2}{s^2+2\zeta\omega_n s+\omega_n} \quad (3)$$

where;
ζ: Damping ratio.
ω_n: Natural frequency (Hz).

III. CIRCUIT DESIGN AND SIMULATION

A. *Phase Frequency Detector / Charge Pump/ Loop Filter*

PFD is a circuit that measures the phase and frequency difference between the signal that comes from the VCO and the reference signal. Outputs pulses (Down & Up) have widths proportional to the phase error. Figure 5 presents a PFD using 2 True Single-Phase-Clock (TSPC) D-Flip-Flop [6]. Clocks are exploited for resetting; F_{ref} to reset F_{fb} and vice versa. In comparison to the conventional PFD, the AND gate, generating the reset signal [7], is removed. The delay due to the propagation of the signal will be saved. The CP converts the error in phase to current information. It has three states as illustrated in Table I.

TABLE I. THREE STATES OF CHARGE PUMP

$\overline{\text{Down}}$	Up	Vctrl
1	0	rise
0	1	reduce
1/0	1/0	unchanged

The main feature of the architecture shown in Figure 6 is the reduction of dead zone (the delay needed on CP digital command signals to avoid all ones condition) and the guarantee of a safe operation without any CP current glitches on the Loop Filter (LF) capacitance. PFD design is suitable for fast PLL operating at higher frequency; we should guarantee that the maximum delay of the PFD is greater than the CP switching speed in order to target a zero PFD dead zone.

B. *Voltage Controlled Osicllator*

The main part of the system is the VCO, based on ring oscillator [8], and designed by a ring connection between five inverters [9]. The charging and discharging currents are equal, both depend linearly on V_{ctrl}. The F_{osc} is proportional to both currents. Thus, F_{out} of the presented architecture varied linearly withV_{ctrl}.

$$F_{out} = \frac{1}{2N \times T_d} \quad (4)$$

Where T_d is the delay introduced by one inverter.

$$T_d = \frac{C_L \times V_{supply}}{I_{cha}} + \frac{C_L \times V_{supply}}{I_{dis}} = \frac{C_L}{I} \quad (5)$$

Where;
V_{supply}: is the supply voltage equal 0.5 V.
C_L: is the load capacitance seen by one gate.
$I_{cha/dis}$: charging and discharging current.

From the presented architecture, $I_{char} = I_{dis} = I$. From (4) and (5), we can deduce that F_{out} is proportional to the current.

From Figure 7, the I_{dis} controlled by an input voltage applied to the gate.

$$I \propto ((V_{ctrl} - V_{th})V_{dsat} - \frac{1}{2}V_{dsat}^2) \quad (6)$$

Where
V_{th} : is the transistor threshold voltage.
V_{dsat} : is the drain source saturation voltage.
From (4), (5) and (6) results:

$$F_{out} \propto \frac{((V_{ctrl} - V_{th})V_{dsat} - \frac{1}{2}V_{dsat}^2)}{2N \times C_L} \propto V_{ctrl} \quad (7)$$

Two series of inverters are used as a buffer to protect the oscillation. When the VCO is close to its steady state, the phase and frequency of the output clock are adjusted slightly to match the phase and frequency of the input reference clock.

C. Feedback Divider

The divider divides VCO's clock to generate FBCLK to compare both phases.

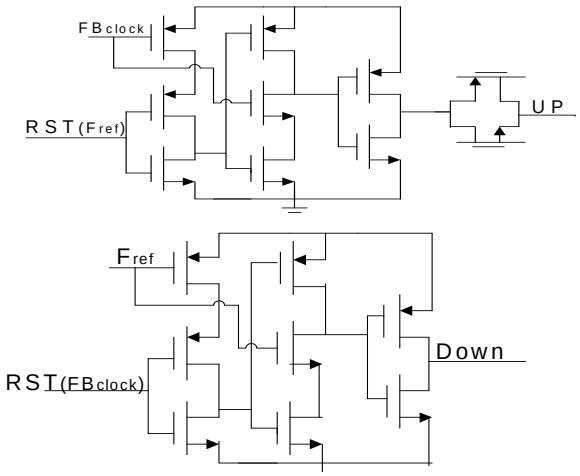


Figure 5. Phase Frequency Detector

It is made up of TSPC Flip Flop [10].

D. Simulation Results

Once the lock is achieved, as shown in Figure 10, after settling time 4µs, shown in Fig (8), the output clock would be equal to $N \times F_{ref}$.

Figure 9 shows the corresponding voltage on the loop filter capacitor as it will saturate when the output clock matches the reference clock frequency.

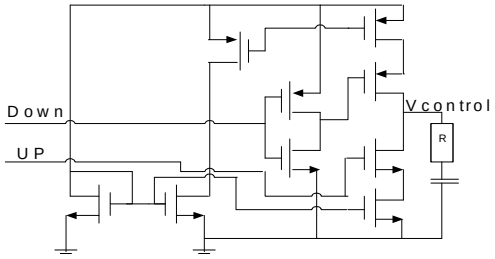


Figure 6. Charge Pump/ Loop Filter

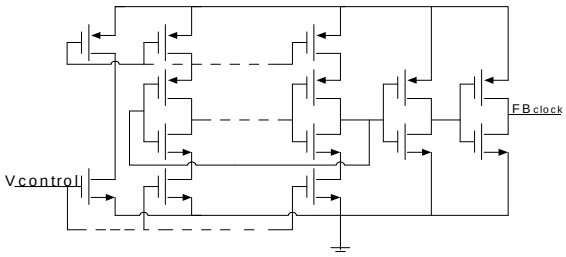


Figure 7. Voltage Controlled Oscillator

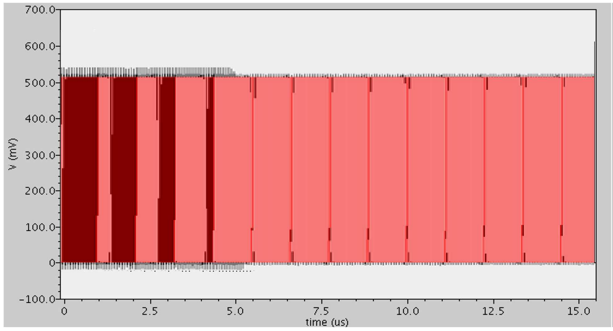


Figure 8. Transient simulation of input/output frequency, N=1

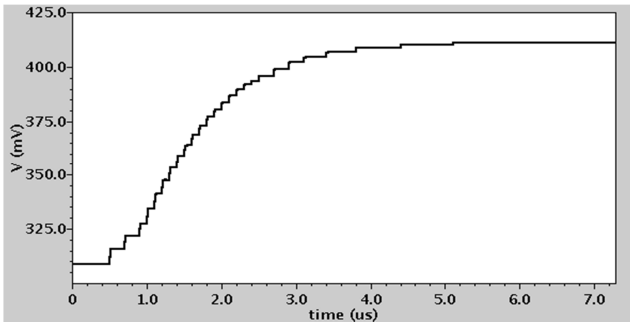


Figure 9. Evolution of voltage control

The controlled voltage is stabilized at 410 mV.

Table II gives a summary of the loop parameters in comparison with other solution.

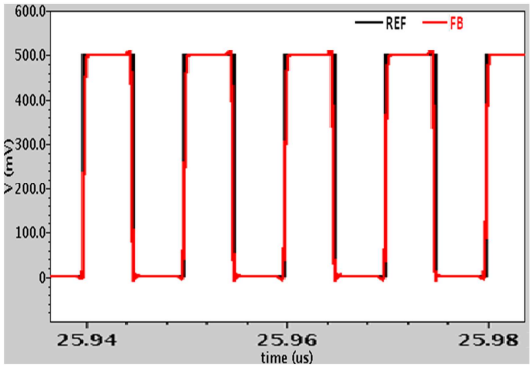


Figure 10.PLL in locked condition reference signal versus feedback signal

TABLE II. LOOP PARAMETERS FOR FIRST ARCHITECTURE

Loop parameter	This work	[11]
F_{osc}	100MHz	3MHz
K_{vco}/K_{vco}(assuming same control current)	303 MHz/V	300 MHz /V
I_{cp}	1.1µA	0.9 µA
C_p	3.61PF	10PF
C_2	--	2PF
PM	83.9°	45.4°
Power consumption	0.375PW/Hz	4.47PW/Hz
Techno	180 nm	90 nm

IV. CONCLUSIONAND FUTURE WORK

The paper presented a designed PLL with low power consumption and small loop capacitance. PFD uses reference and feedback clock to reset each other; UP and DOWN signals don't get high simultaneously, yield to low glitches for a small loop capacitance value, which is the gain as we kept low area required for the application.

CP current is reduced given that its phase noise transfer function is a LPF (decay in the band of interest) while a VCO large current to reduce its output phase noise given that its phase noise transfer function is a HPF. System and design are done to optimize the power consumption. System stability is guaranteed as PM >60°.

In our future work, we target designing this PLL to operate at higher frequency; for that reason, we have discussed above about the condition for zero PFD dead zone.

ACKNOWLEDGMENT

The author would like to express her sincere appreciation to her advisor Prof. El-Sayed Eid for his enthusiastic guidance and advice throughout this work and Mohamed Saleh for his support on technical aspects.

REFERENCES

[1] B. DeMuer and M. Steyaert, "CMOS fractional-N synthesizers design for high high spectral purity and monolithic integration," Springer, 2003.

[2] Y. Guo and Z. Xie, "Design of PLL Frequency Synthesizer in Frequency Hopping Communication System," Communications and Mobile Computing (CMC), 2010 International Conference on, vol. 3, pp. 138-141, 12-14 April 2010

[3] N. M. H. Ismail and M. Othman, "Low Power Phase Locked Loop Frequency Synthesizer for 2.4 GHz Band Zigbee," American Journal of Engineering and Applied Sciences, vol. 2, 2009, pp. 337-343.

[4] D. Duarte, N. Vijaykrishnan and M. J. Irwin, "A complete phase-locked loop power consumption model," Design, Automation and Test in Europe Conference and Exhibition. Proceedings, 2002, doi: 10.1109/DATE.2002.998464.

[5] W. L. Brogan, , G. K. F. Lee, A. P. Sage, B. C. Kuo, , C. L. Phillips, R. D. Harbor, R. G. Jacquot, J. E. McInroy, D. P. Atherton, J. S. Bay, W. T. Baumann, M. Y. Chow, "ControlSystems," The Electrical Engineering Handbook Ed. Richard C. Dorf Boca Raton: CRC Press LLC, 2000.

[6] D. R. Sulaiman, "A902-928 MHz PLL Design and Simulation for Biomedical Applications," The 5th International Conference on Information Technology, 2011.

[7] S. Edgar and S. Sinencio, "CMOS PLL Synthesizers: Analysis and Design," Springer, 2005.

[8] B. Razavi, "Design of monolithic phase locked loops and clock recovery circuits-A tutorial," Monolithic Phase Locked Loops and Clock Recovery Circuits-Theory and Design, IEEE Press, 1996.

[9] B. Razavi and R. Behzad, RF microelectronics, Prentice Hall PTR Upper Saddle River, NJ, 1998.

[10] J. A. McNeill and D. Ricketts, The Designer's Guide to Low Jitter Oscillators, Springer Verlag, 2009.

[11] S. H. Yang, Y. You, and K. R. Cho, "A New Dynamic D-Flip-Flop Aiming at Glitch and Charge Sharing Free," IEICE Transactions on Electronics, vol. 86, 2003, pp. 496–505.

[12] H. G. Ozsema, "Ultra-Low-Power and Widely Tunable PLL," Master Thesis, EPFL, January 2010.

 ISBN: 978-1-61208-317-9

Studying Risks in Space Mission Communications

Objectives, approach and expected results

Julio Vivero
Consulting Area
GMV
Barcelona, Spain
jvivero@gmv.com

Luca del Monte
Senior Strategy Coordinator
European Space Agency
Paris, France
luca.del.monte@esa.int

***Abstract*—The paper presents the study on cybersecurity risks for space missions launched by the European Space Agency. This one-year study covers all lifecycle phases within civilian space missions and includes all their actors and elements both at the ground and the space. However, the paper focuses in how mission communications are addressed in the study, the approach followed to identify their risks and the expected results.**

Keywords-Risk; safeguards; communications; missions; taxonomy

I. INTRODUCTION

Cybersecurity is a growing concern for governments, public and private entities, and citizens [1][2]. Our reliance on information systems and technology for most of our daily activities is a reality. This fact converts Information Technology and Communications (ITC) into a more attractive target for strategic advantage, espionage, fraud, notoriety, etc.

Space missions are not an exception to this general trend [3]. Nowadays, space missions provide valuable services to society in many fields: from navigation, to earth observation, weather forecasting or communication, sometimes even providing significant revenues to the operators. As such, cyber threats to space missions will continue to grow in the near future [4].

The European Space Agency (ESA) has perceived this trend in the last years and has launched a study to analyze cybersecurity risks to space missions and recommend remediation safeguards. The goals of the study are providing tools to mission planners in order to introduce adequate levels of security in missions in an easier and more efficient way and promote awareness on the increasing cybersecurity risks. The study assesses risks during all phases in the lifecycle of space missions [5] and on all elements and actors involved in the mission.

As crucial element in most, if not all, space missions, the communication links and equipment are carefully addressed in the study. Furthermore, ground-space communications are one particularly critical element in space missions due to their exposure as it is based in an open and easily accessible physical media: the radio channel.

Although the study is broad and generic in nature, this paper focuses in the methodology that will be followed to identify, categorize, and mitigate risks on space mission communications.

In Section II, the study main objectives and how they will be addressed will be described. The methodology followed to identify, assess, and manage risks on communication elements will be explained in Section III. The expected outcomes and impact of the study will be defined in Section IV. Finally, Section V summarizes the main conclusions, future work, and milestones of the study.

II. OBJECTIVES

The main goal of the study is "to support ESA in the establishment of technical recommendations and a policy" that help mission planners determine which security safeguards shall be implemented within each particular phase of their mission lifecycle.

It is understood that the result of the study is expected to be an important tool for mission planners. With this tool mission planners shall be able to easily understand the threats, vulnerabilities, and resulting risks that apply to their specific mission category, including all communications issues, and more importantly, safeguards they should implement to reduce that risk to an acceptable level.

The concept behind the study is, given the similarities in missions from the same category in the taxonomy developing, a-priori, a risk assessment and risk treatment plan for each category so that it can be reused in all subsequent missions. Tailoring of the results with mission particularities, such as the relative value of the mission (when compared with its peers in the same mission category) will be possible.

Another complementary objective is to "raise awareness in the space community about the cyber-security issues". In this sense, both the final report and the executive summary will present the information in a way that is easily readable, understandable, and usable by a broad audience in the field of space operations.

To accomplish the above objectives a new risk assessment methodology has been designed specifically for the study. The following section describes in detail this methodology focusing on how space mission communication aspects are addressed.

 ISBN: 978-1-61208-317-9

III. METHODOLOGY

A. Principles

The proposed approach leverages IT-Grundschutz [6] risk assessment methodology, defined by the German Federal Office for Information Security, and adapts it to the particular needs of the study. IT-Grundschutz is based on the modules concept. Modules are representations of threats, safeguards, and ultimately risks for technological components that can be easily applied by organizations saving the effort of performing risk assessments for these technological components since these assessments are done beforehand and reflected in the modules.

Our methodology follows this concept and creates modules for each mission category and lifecycle phase which can be tailored in a number of ways to fit in mission particularities. This approach provides some benefits in respect to other risk assessment and risk management approaches, namely the possibility to customize and tailor the risk assessment to the concrete particularities of the space mission and the ease of use from the mission planner perspective.

In addition to this modularity and customization principles, the methodology is also systematic, to guarantee that no relevant aspect is left aside. Efficiency is essential to optimize the effort required to produce the modules. Commonalities of space missions, even in different categories, shall be exploited to minimize the effort required. Finally but no less important, the methodology and its instantiation shall be well documented to provide evidences and justifications for decisions taken and to guarantee that results are consistent among methodology instantiations.

B. Phases

Fig.1 shows the methodology phases and their interactions.

The first phase is the taxonomy definition. This is a crucial activity in the study since a great percentage of the results applicability and study success depends on how missions are categorized. All missions falling with the same category shall share attack motivations for attackers, threats, mission elements, vulnerabilities, resultant risks. Also, identified safeguards shall be applicable to all of them. There shall not be too many categories because the effort efficiency principle would be lost, neither too few because the risk of results not being applicable to all missions that fall within that category.

There are several aspects which are security relevant but do not depend on the mission category, but rather on business models, or mission specific design decisions. Examples of these aspects are subcontracting of mission facilities, sites or even operations, or more relevant to the current paper, the type of communications present in the mission: ground-space, ground-ground or space-space. To address these aspects additional packages will be defined to complement modules with the threats, vulnerabilities, risks, and safeguards linked with them. Hence, the mission planner will select the main mission module applicable for its mission category plus the additional packages applicable to that mission.

Ground-space communications is considered applicable to almost all space missions and hence will be included in the mission category modules. However, ground-ground and space-space communication packages will be developed covering all the risk relevant elements for this kind of communications.

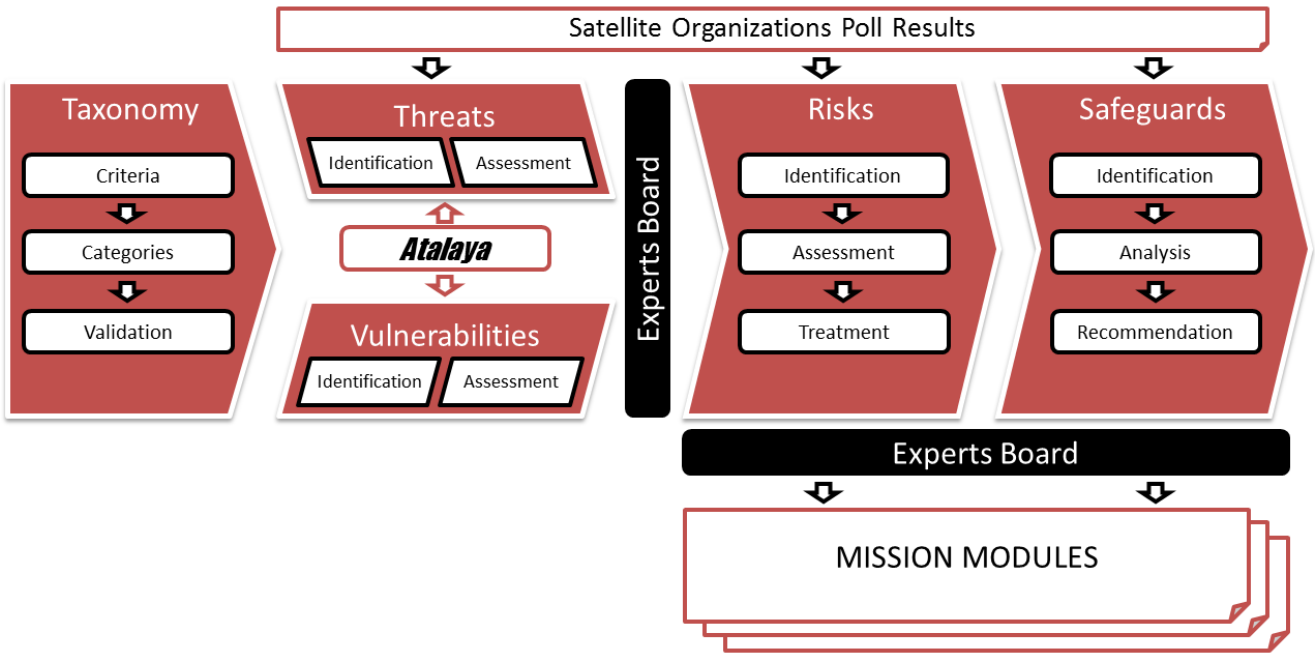


Figure 1. Cybersecurity Study Methodology.

 ISBN: 978-1-61208-317-9

The second phase in the study is the threats assessment for each lifecycle phase of each mission category. The identification of threats will be developed based on threat catalogues (such as IT-Grundschutz catalogue [7], ISO27005 threats catalogue [8] or CCSDS 350.1-G-1 [9]), knowledge of the study team members and the results of a poll among different satellite organizations. These threats, and especially those affecting communications, shall be tailored to space mission particularities. Aspects, such as jamming, protocol efficiency, lag, and space and earth weather shall be taken into account. Jamming is clearly the main threat in relation with space-ground communications but there are other threats which shall not be neglected neither, e.g. eavesdropping, replay attacks or even man in the middle types of attack.

Vulnerabilities will be identified for each asset at each lifecycle phase of the space mission categories. Common weaknesses, insecure configurations, safeguard gaps or implementation flaws will be analyzed as sources of vulnerabilities. The lack of encryption and authentication in messages, weak cryptographic key management practices, single points of failure in communication channels, lack of Denial of Service (DoS) resilience, inexistent or insufficient anti-jamming safeguards and other issues will be examples of vulnerabilities considered from the communications standpoint.

GMV Atalaya service will support the previous phases by providing evidences that justify threat and vulnerabilities values. Atalaya monitors the internet to identify potential attack information, data of interest to attackers or vulnerabilities evidences.

When a threat has been identified with the potential to exploit existing mission vulnerabilities a risk is derived. Resulting risks are assessed based on safeguards traditionally implemented in missions, and then aggregated and prioritized.

During the last phase of the methodology, a set of safeguards will be recommended to mitigate identified risks. As with threats, safeguards will be obtained from catalogues [7][9][10][11], study team knowledge and the satellite organizations poll. Several safeguard implementation alternatives will be proposed, each with a cost-benefit analysis and a description of the remaining residual risk. Also, safeguards will be classified in basic safeguards (recommended for all kind of missions within the category) or high protection safeguards which would be applicable for those missions with higher protection requirements. In some cases areas of further research will be pointed out for investigating more efficient safeguards for particular threats, e.g. innovative anti-jamming techniques, new telemetry, tracking and commanding (TT&C) approaches, etc.

IV. EXPECTED RESULTS

The main results of the study are the modules and additional packages. Each module will contain the risk assessment and risk treatment result for a lifecycle phase of a mission category.

Contents of the modules include the list of applicable threats for each asset and actor, the vulnerabilities, the associated risks, and the recommended safeguards.

Additional packages present the same contents as modules but focused in aspects which are specific of each mission and cannot be extrapolated from the mission category. It is planned to produced additional packages for space-space communications, ground-ground communications, multi-organizational missions, outsourced mission elements, missions with high public visibility and for the type of spacecraft bus used (serial or ad-hoc).

Mission planners will then need to select the appropriate module and complement it, as needed, with the additional packages to get the list of recommended safeguards to include in the mission so that cyber-security risks are mitigated.

The mission taxonomy elaborated during the first phase of the study is another interesting result by itself. It provides a classification of space missions per their cyber-security properties.

At this date, the mission taxonomy phase of the study has been concluded. The following mission categories have been identified:

- Launchers
- Space Tourism
- Manned In-Orbit Infrastructure / Planetary Exploration
- Supply and Re-Entry Vehicles.
- Navigation
- Communications
- Weather Forecast
- Earth Observation Mapping
- Earth Observation Surveillance
- Earth Observation Environmental Monitoring
- In-Orbit Servicing
- Scientific – Up to GEO Orbit
- Scientific – Above GEO Orbit
- Space Situational Awareness (SSA)

Fig.2 below shows the percentage of past, present, and future ESA missions that fall within each mission category according to public information at ESA website (www.esa.int/ESA/Our_Missions).

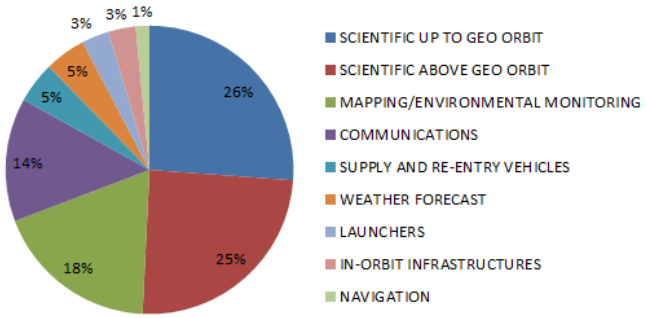


Figure 2. Percentage of ESA missions per category

The above deliverables are clearly presented, described, and explained in a final report and executive summary. These reports will summarize all information in a way to be

 ISBN: 978-1-61208-317-9

easily readable, understandable, and usable by a broad audience in the field of space operations

One element included in the methodology, and relevant to the expected results, is the Experts Board concept. To guarantee the quality, applicability, and accuracy of all results, an Experts Board formed by highly qualified and experienced GMV personnel has been formed to review all study deliverables before they are published. In this way, threats, vulnerabilities, risks, and safeguards will be fully useful and applicable to mission planners.

V. CONCLUSIONS AND FUTURE WORK

GMV and ESA are convinced of the potential benefits which can be generated by the study results. Namely, increasing awareness in cyber-security among space mission stakeholders and actors, enhancing the security of current and future missions and thus mitigating their cyber-security risks and finally reducing the effort and knowledge required for mission planners to decide what safeguards shall be implemented in the mission and how.

These benefits will only be achieved if certain conditions are given. First, and most important, the results shall be very easy and intuitive in their exploitation by mission planners. That is, the effort to learn how to use them and apply them shall be as low as possible. Second, risk modules shall be fully applicable to every mission according to its category. If a high percentage of threats, vulnerabilities and safeguards within a module are clearly not applicable to the mission; then mission planners will be discouraged and discard its use. Third and last, recommended safeguards included within the modules shall be effective in the mitigation of space mission risks. Otherwise, the study results would be either complemented or replaced with a different approach.

Looking further into the future it is also important to consider that results shall be simple to maintain and easily updateable with new modules and additional packages to accommodate new mission categories or aspects to consider in terms of cybersecurity. To facilitate these maintenance tasks the methodology shall be described in detail, including guidelines on how to implement the methodology, criteria to consider and rationale followed.

The study started in July 2013 and is expected to be executed during one year, until July 2014.

REFERENCES

[1] European Commission, "EU Cybersecurity plan to protect open internet and online freedom and opportunity – Cyber Security strategy and Proposal for a Directive," Digital Agenda for Europe, February 2013

[2] National Institute of Standards and Technology (NIST), "Cybersecurity Framework," Information Technology Laboratory, February 2013.

[3] Staff Writers, "The State of NASA's Cybersecurity," Space Safety Magazine, March 2012.

[4] Kris Osborn, "Air Force Faces Increasing Space Threats: Shelton," DefenseTech, September 2013.

[5] European Coordination for Space Standardization, "Project planning and implementation," ECSS-M-ST-10C rev1, 3rd issue rev. 1, March 2009.

[6] German Federal Office for Information Security, "Methodology for evaluating usage and comparison of Risk Assessment and Risk Management items," Deliverable 2 version 1, April 2007.

[7] German Federal Office for Information Security, "Reference source for threats, vulnerabilities, impacts and controls in IT risk assessment and risk management," Deliverable 3 version 1, April 2007.

[8] International Standards Organization, "Information technology — Security techniques — Information security risk management," ISO/IEC 27005, 1st ed., June 2008.

[9] Consultative Committee for Space Data Systems, "Security Threats Against Space Missions," CCSDS 350.1-G-1, October 2006.

[10] National Institute of Standards and Technology, "Recommended Security Controls for Federal Information Systems and Organizations," SP800-53 rev.3, August 2009.

[11] International Standards Organization, "Information technology - Security techniques - Code of practice for information security management," ISO/IEC 27002:2005, 2nd ed., June 2005.

 ISBN: 978-1-61208-317-9

Telemetry Storage and Donwlink Management for a LEO Satellite

Dongseok Chae
Satellite Flight Software Department
Korea Aerospace Research Institute
Daejeon, Korea
dschae@kari.re.kr

***Abstract*—Because Low Earth Orbit (LEO) satellite has very limited contact time between satellite and ground station, all telemetry data generated on satellite are stored in a mass memory and downlinked to the ground together with real-time data during the contact time. There are two mass memory modules. Normally, one mass memory module is used and the other is ready for backup. If required, both memory modules can be used at the same time. If the selected module is not available, the other module can be automatically used for telemetry storage without ground intervention. Two downlink channels are available at the same time but only one channel should be used for downlink. The downlink channel can be changed automatically by on-board flight software as well as step by step by ground command. This paper presents the telemetry storage and downlink management method implemented in flight software of a LEO satellite developed in Korea. All functions related to telemetry storage and downlink management were fully verified through the real satellite operation as well as the various tests during satellite development phase.**

Keywords-telemetry; downlink; flight software.

I. INTRODUCTION

On-Board Computer (OBC) for a LEO satellite developed in Korea Aerospace Research Institute (KARI) is one unit box which includes primary and redundant side. It consists of Processor Module (PM) for supplying Flight Software (FSW) execution platform, Tele-command and Telemetry Module (TCTM) for performing uplink and downlink, Bus I/O Module (BIOM) for interfacing with bus units, Payload I/O Module (PIOM) for interfacing with payload units, power converter and the others [1][2]. Figure 1 shows the telemetry data downlink interfaces.

TCTM modules are cross-strapped with two processor modules and activated simultaneously as hot redundancy. Mass Memory Management Unit (MMU) receives telemetry frames from PM through space-wire interface [3] and transmits them to Telemetry Unit (TMU) or stores into mass memory and reads the stored frames from mass memory and transmits them to TMU. TMU receives the telemetry frames from MMU and transmits them to transponder. The mass memory for telemetry storage consists of two modules. TCTM has its own mass memory. PM can access each mass memory, but MMU can access only its own mass memory.

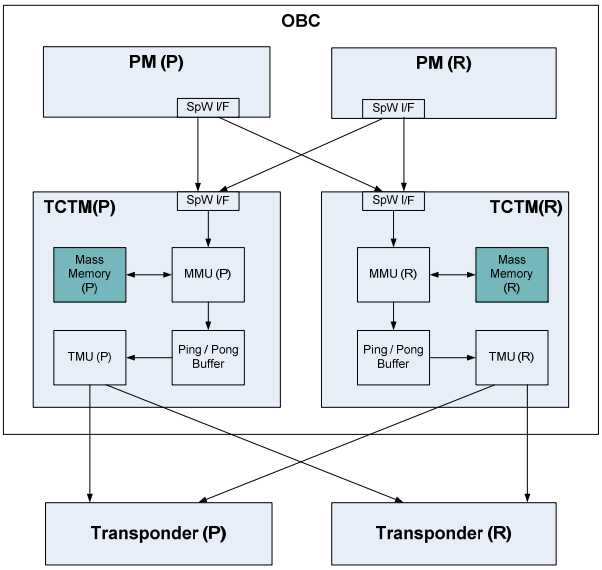


Figure 1. Telemetry Data Downlink Interfaces

Telemetry uses Consultative Committee for Space Data Systems (CCSDS) compliant VCDU frames [4][5]. Telemetry data storage method is described in section 2. Downlink management method is described in section 3.

II. TELEMETRY DATA STORAGE

A. Mass Memory Data Structure

Figure 2 shows the mass memory data structure. The mass memory for telemetry data storage is divided into blocks that are called pages. Each memory module has 2048 pages and a page consists of 128 Kbytes. Normally, one mass memory module is used and the other is ready for backup. If required, both memory modules can be used at the same time. If the selected module is not available by memory full condition or interface problem, the other module can be automatically used for telemetry storage without ground intervention. The page structures are shown in Figure 3. 585 Virtual Channel Data Unit (VCDU) frames are stored on each page.

 ISBN: 978-1-61208-317-9

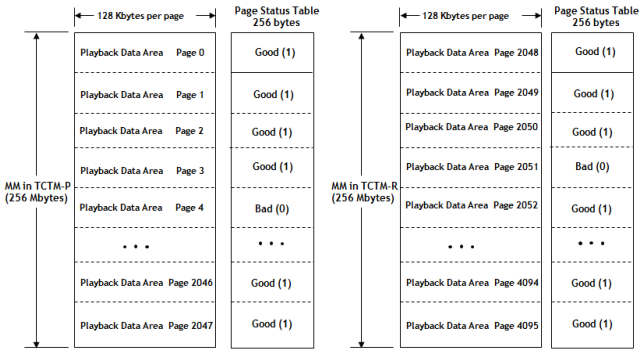


Figure 2. Mass memory structure for telemetry storage

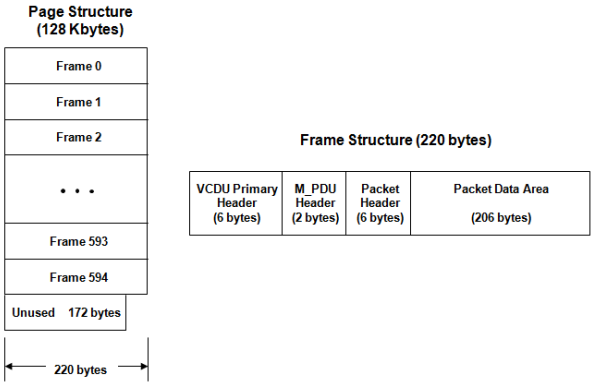


Figure 3. Page structure of mass memory

B. *Telemetry Data Type and Storage*

All telemetry data are formatted into Consultative Committee for Space Data Systems (CCSDS) compliant VCDU frames [4], and stored into mass memory. Telemetry types are shown in Table I.

TABLE I. TELEMETRY TYPE

Type		Max # of Frame	Period (sec)
Real Time	SOH	4	1
Storage-Only	SOH	5	1
	POD	1	1
	PAD	4	1
	PLD SOH 1	5	1
	PLD SOH 2	4	24
	PLD SOH 3	3	24
	PLD SOH 4	3	64

Each telemetry frame has its own VCDU ID and sequence counter, and stored in satellite memory in chronological order without overwriting older data which has not yet been downlinked. Newly generated telemetry frames are stored on the memory location pointed out by current write pointer. If the mass memory page status is marked as bad, the page is not used anymore for telemetry storage. If the selected module is not available by memory full condition or interface problem, the other module can be automatically used for telemetry storage without ground intervention. In case of real-time telemetry, they are stored into mass memory and downlinked to the ground simultaneously. Two real-time SOH frames in low downlink rate, or two to four real-time frames in high downlink rate are generated, stored into mass memory and downlinked to the ground every second. In case of storage only telemetry, they can be generated and stored into mass memory regardless of the hardware downlink rate.

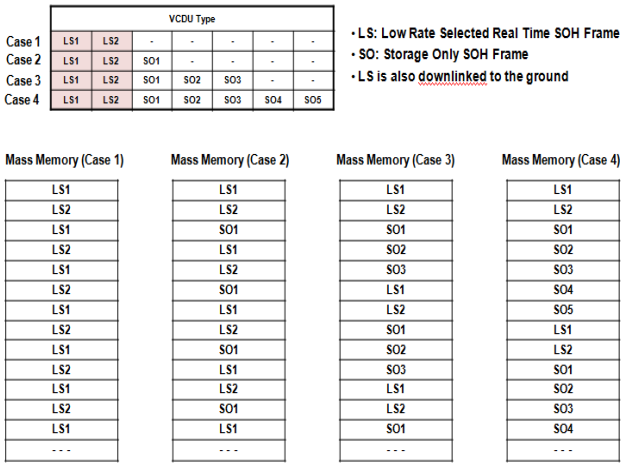


Figure 4. Telemetry storage example (low rate)

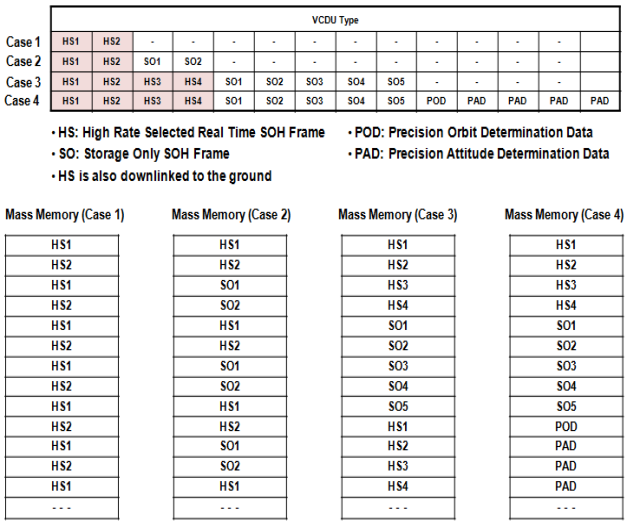


Figure 5. Telemetry storage example (high rate)

Figures 4 and 5 show the examples of several cases of telemetry storage in the mass memory in accordance with the hardware downlink rate.

 ISBN: 978-1-61208-317-9

III. DOWNLINK MANAGEMENT

A. Downlink Rate

There are two downlink rates, low downlink rate (4096bps) and high downlink rate (1.5626Mbps). Only real-time telemetry frames can be downlinked to the ground in low downlink rate. In high downlink rate, there are two downlink modes, real-time mode and playback mode. Only real-time telemetry frames are downlinked in real-time mode and real-time and stored telemetry frames in the mass memory are downlinked in playback mode. Two real-time frames in low downlink rate and two to four real-time frames in high downlink rate can be downlinked to the ground every second. Total 762 telemetry frames including two to four real-time frames can be downlinked to the ground every second in high downlink rate. Figure 6 shows the downlink examples in high downlink rate.

	Downlink Frame Sequence				
	1	2	3	4	5~762
Case 1	HS1	HS2	PB	PB	PB
Case 2	HS1	HS2	HS3	PB	PB
Case 3	HS1	HS2	HS3	HS4	PB
Case 4	HS1	HS2	HS3	HS4	Fill

- **HS : High Rate Selected Real Time SOH Frame, PB : Playback Frame**
- **762 frame/sec:**
  - Case 1: 2 RT + 760 PB frames (Playback Mode)
  - Case 2: 3 RT + 759 PB frames (Playback Mode)
  - Case 3: 4 RT + 758 PB frames (Playback Mode)
  - Case 4: 4 RT + Fill Patterns (Real Time Mode)
- **PB is replaced by fill pattern if PB mode is not selected**

Figure 6. Downlink example

B. Downlink Operation

Playback operation is started by ground command. Several types of playback command are used. Telemetry frames are downlinked from the last playback pointer, and proceed until the current write pointer is reached upon receipt of the PB Start command. Telemetry frames are downlinked from the specified location, and proceed until the current write pointer is reached upon receipt of the PB Location command. All Telemetry frames are downlinked upon receipt of the PB All command. Telemetry frames are downlinked from a start address to the end address upon receipt of the PB Specified Area command. The downlink mode is automatically switched to the real-time mode when playback is completed.

C. Downlink Channel and Rate control

Because two TCTM modules are operated as hot redundancy, two downlink channels are available at the same time but only one channel should be used for downlink. Each TCTM has own mass memory and is able to access only own mass memory. In case of real-time mode, any downlink channel can be used because it does not need to access mass memory. But, in case of playback mode, the downlink channel of the TCTM including the selected mass memory should be used. If the mass memory is changed to the other module, the downlink channel should be changed in accordance with the mass memory change. In this case, the downlink channel can be changed automatically by on-board flight software as well as step by step by ground command. If two mass memory modules are used at the same time, change between two mass memory modules and two downlink channels are performed automatically by on board flight software for performing playback operation continuously.

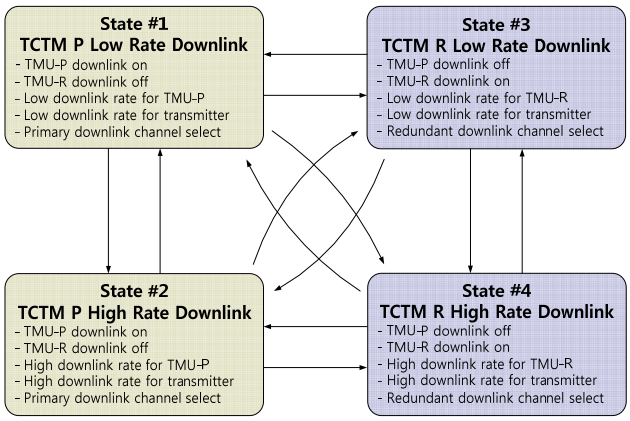


Figure 7. Downlink channel and rate state transition diagram

Figure 7 shows the downlink channel and rate state transition diagram. There are 4 states and 12 state transition cases to change the downlink rate and the downlink channel.

IV. CONCLUSION

Telemetry storage and downlink management method implemented in flight software of a LEO satellite developed in Korea were described in this paper. In case all types of telemetry frame are selected, it takes about one day to fill up one mass memory module. It takes about 27 minutes to play back all area of one mass memory. Just one ground station is enough to receive the telemetry.

Most of all functions required to manage mass memory, to store telemetry and to downlink telemetry data are automatically performed by on-board flight software. Operator has only to select the telemetry types to be downlinked to the ground or stored into mass memory, and to send command to start playback. If the mass memory has interface problem or fault in specific area, it is possible to change to other module and to initialize all or specific area of mass memory by ground command. All functions related to telemetry storage and downlink were fully verified through the real satellite operation as well as the various tests during satellite development phase.

 ISBN: 978-1-61208-317-9

REFERENCES

[1] Y. K. Lee, J. H. Won, and S. K. Lee, "A study of common OBC HW architecture for LEO & GEO satellite", IEICE Technical Report, Oct. 2009, pp. 77-83

[2] D. Y. Kim, K. H. Kwon, J. W. Choi, J. I. Lee, H. J. Kim, "Design of a new on-board computer for the new KOMPSAT-bus", IEEE Aerospace conference, Mar. 2005

[3] ATMEL, SMCS332 User Manual, DIPSAP-II Consortium, 1999.

[4] CCSDS 102.0-B-5 Consultative Committee for Space Data Systems Packet Telemetry

[5] CCSDS 873.0-R-2, Spacedraft onboard interface services—file and packet store services, 2011, Draft recommendation for space data system practices

[6] J. R. Wertz, W. J. Larson, Space Mission Analysis and Design, Third edition, Microcosm Press, Torrance Ca, USA, 1999

[7] G. Casperson, Software System Development for Spacecraft Data Handling & Control, TERMA Electronics AS, 1999.

 ISBN: 978-1-61208-317-9

A Maturity Model for Large and Complex Programs/Projects Management

Haibin Liu
China Aerospace Academy of Aerospace Systems Science and Engineering
University of South Australia
Beijing, P.R.China
e-mail: liuhb@spacechina.com

Gongtao Wang
China Aerospace Academy of Aerospace Systems Science and Engineering
Beijing, P.R.China
e-mail: wanggt@spacechina.com

Jiaming Liu
China Aerospace Academy of Aerospace Systems Science and Engineering
Beijing, P.R.China
e-mail: JaminLiu1988@gmail.com

***Abstract*—In this paper, the features of the national key science and technology programs of China are analyzed. The related management maturity models studied abroad are also introduced. The weights of each management factor in the execution process of the national key science and technology programs are calculated through the combination of Analytic Hierarchy Process and Entropy Method. Therefore, a management maturity model which can be suitable to the national key science and technology programs of China is proposed. China High resolution Earth Observation System (CHEOS), which is one of 16 national key science and technology programs of China, is selected as a study case. By utilizing Fuzzy Comprehensive Evaluation (FCE), the management maturity level of CHEOS is assessed. The assessment result can provide an aid to the decision-making of the CHEOS management. Furthermore, a new management method for the national key science and technology programs of China is advanced.**

Keywords-national key science and technology program of China; program/project management; maturity model; analytic hierarchy process; entropy method; fuzzy comprehensive evaluation

I. INTRODUCTION

A. *Features of the national key science and technology programs of China*

To accomplish the nation goals, the national key science and technology programs of China are to form great strategic product, key technology and/or major project through breakthrough in core technology and integration of resource under specific time limits. Therefore, each key science and technology program is a large and complex system. In summary, there are 6 main features of the China national key science and technology programs [1][2][3]:

(1) Government dominance: The national key science and technology program is an important organization form to fulfill the national strategic demands. The implementation of these projects must serve the national goals and reflect the national interests, i.e., the government must arrange the strategic plans and funding allocations under consideration of the national interests and future strategies. Thus, it is the government that supports a main area till the realization of a local leap-forward development and drives the breakthroughs in strategic industries.

(2) Resource integration: The formation of the great strategic product and key common technology comes from the comprehensive integration of different subjects and/or various areas. That is to say, the implementation of the national key science and technology programs demand the information opening, the resource sharing and the collaborative innovation among every participating department.

(3) Management complexity: Due to the limitation of the power and resource of a single existing organization, the interdisciplinary and interdepartmental features of the program are hard to overcome. Furthermore, during the implementation of the program, there are numbers of stake-holders and funding resources, which are much beyond the prowess of normal project management.

 ISBN: 978-1-61208-317-9

(4) Investment enormity: The implementation of each national key science and technology program demands enormous amount of investment.

(5) High risk: There are different kinds of risks e.g., the technology risk, the policy risk and the market risk etc involved in the national key science and technology programs. Therefore, the implementation of the program is under influence of many factors, all degrees of uncertainty and risks.

(6) Influence universality: The national key science and technology programs are to accomplish national goals and to meet the national demands. The results of technology research and application coming from these programs are the source of the original innovation and the proprietary intellectual property, hence reflect the core competitiveness of China. Ultimately, the outcome of the program can form a deciding influence in the related area and have a wide influential effect on the whole economy of the nation.

B. Current project management maturity models

1) The concept and significance of project management maturity model

The project management maturity model can express the management ability of an organization, and the model can also help to implement the project according to the scheduled goal and condition.

The project management maturity model provides an assessment method and an improvement framework for project management as a new idea. Based on the process of project management, the organization can range the level of project management from chaos to order and forms an upgrade platform step by step. The current management level is the basic of the higher management level. The upgrade of the ability of the project management maturity is also the accumulation of project management level. With the help of the project management maturity model, the organization can find the defects of the project management and recognizes the weak link of the project management and then improves the project management ability steadily.

2) Organizational project management maturity model

Organizational project management maturity model (OPM3) is a new standard which was released by Project Management Institute in 2003 [4].

To improve project management ability of an organization, and associate the project with organization strategy tightly, OPM3 provides the method. To confirm the state of an organization and formulate plans, OPM3 supplies plenty of knowledge for users to acknowledge project management and provides a standard to assess the project management ability.

OPM3 consists of three dimensions, the first dimension presents maturity levels, the second dimension presents project management areas and project management processes, the third dimension presents three domains of the project management. The maturity level includes Standardizing, Measuring, Controlling and Continuously Improving. The project management area includes Project Integration Management, Project Scope Management, Project Timeline Management, Project Cost Management, Project Quality Management, Project Human Resource Management, Project Communication Management, Project Risk Management and Project Purchasing Management. The project management process includes Initiating Process Group, Planning Process Group, Exceuting Process Group, Monitioring & Controlling Process Group and Closing Process Group. The domain of the project management includes Project Management, Program Management and Portfolio Management.

3) K-PMMM

K-PMMM was advanced by Kerzner [5] in 2001, this model is mainly used by enterprise, and the level of the project management maturity is Common Language, Common Processes, Singular Methodology, Benchmarking and Continuous Improvement.

The design of the questionnaire for K-PMMM is different from other models. The questionnaire provides several objective self-assessment questions in different aspects. In each level, there are different kinds of questions, e.g., there are 80 choice questions in the first level. Through analyzing the answer to these questions, organization can find the project management problems of an organization may have and then improves the project management ability.

4) Capability maturity model for software

To fulfill the United State Federal Government's appeal to assess the ability of software suppliers, CMU started to study CMM (Capacity Maturity Model) in 1986 and released the CMM 1.0 in 1991.

CMM has a detailed description for the definition, implementation, measurement and improvement of software. The core of CMM treats software development as a whole process, organization researches the development and maintenance of software to accomplish the commercial goal in a more scientific and standard way. The maturity level of

 ISBN: 978-1-61208-317-9

CMM includes Initial, Repeatable, Defined, Managed and Optimizing. CMM provides a process improvement framework by assimilating experiences from other software development process.

The process of software development includes many activities; these activities contain technology management of software development, quality management, risk management and so on. CMM strives to improve the management ability of software development. The definition of each maturity level is depicted by CMM definitely. CMM supplies a series of approaches, which help the organization to improve the process of software development step by step. So, the ability of software development is improved by CMM gradually.

II. MATURITY MODELS FOR NATIONAL KEY SCIENCE AND TECHNOLOGY PROGRAM OF CHINA

The propose of the maturity model for national key science and technology programs of China is based on CMM; both the ability of software development and the management of national key science and technology program of China are based on process, the maturity model sees activities of software development as an integer, the development of the software is not to create breakthrough technologies solely. It is the same as the management of national key science and technology programs of China.

A. The maturity levels and features of national key science and technology program of China

The partition of each maturity level is to establish the standard for each maturity level, and then the definition for each maturity level is established according to the standard of each maturity level. The divide of each maturity level is based on some significant models. There are 5 levels: Disordered level, Repeatable level, Standardized level, Predictable level, Optimized level, the levels are showed in Figure 1, and the definition of each level is as follows in Table 1:

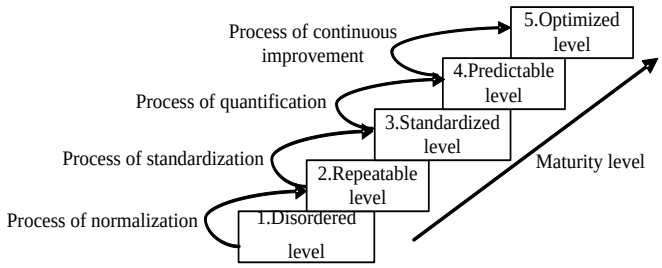


Figure 1. The five levels of maturity model

TABLE I. THE DEFINITON OF EACH MATURITY LEVEL

Level	Definition
Disordered level	The management in this level is disordered; some management factors such as quality, the maturity of technology, investment and risk can not be predicted. The conditions to implement the project are poor, when the problems occur in the process of the project; the organization is likely to give up the plan that has already existed and turns to solve the current problems. And the current problems are solved by the experience of personal; this way can not present the ability of a system.
Repeatable level	The management in this level has an effective organization to manage the project, the management organization of the project forms relative policies about the project and some mature methods to solve the problems. The organization manages these projects according to experience. The management in this level is institutionalization and repeatability.
Standardized level	The management activities in this management level are full, and the standard of the organization is based on these series of activities. The organization makes norm and standard for management plan, quality and so on. There are plenty of useful information can be used and shared, it also becomes the habit of the organization. At the same time, some management factors such as quality, risk and investment can be controlled.
Predictable level	The organization makes quantitative analysis for the project in this management level, in different stage of the project, there has a quality method corresponding to the stage. And some management factors such as quality, risk and investment can be quantified in the process of management. This way can reduce the uncertainty during the process of the project.
Optimized level	The organization optimizes the process of the management continuously, the weaknesses and advantages will be found with different measures to achieve the goal of preventing defects. At the same time, the organization analyses the cost and the benefit of the project, and then some suggestions are proposed according to the analysis. The characteristic of this management level is that the ability of project management can improve continuously by itself in order to prevent the same mistake.

B. The structure of national key science and technology program of China

The content of each dimension is: there are 5 maturity levels in the 1st dimension; it includes Disordered level, Repeatable level, Standardized level, Predictable level, Optimized level, the 2nd dimension is project management evaluation factor system, it includes 10 management areas about national key science and technology program of China, there are 41 second-class targets, the 3rd dimension is the process of project management, it is the life cycle of national key science and technology program of China, it includes Initiating Process Group, Planning Process Group, Executing Process Group, Monitoring & Controlling Process Group, Closing Process Group. The structure of model is shown in Figure 2, the evaluation index system of national key science and technology program of China is shown in Table 2:

 ISBN: 978-1-61208-317-9

TABLE II. THE EVALUATION SYSTEM FOR THE NATIONAL KEY SCIENCE AND TECHNOLOGY PROGRAM OF CHINA

Criteria	Main Factors
Project management	Project approval management
	Project implement and acceptance check management
	Project process management
	Chief engineer and general director management
Project quality control and management	The construction of quality management system
	Quality plan
	Quality control
	Quality assurance
	Quality improvement
Project technology maturity management	Confirmation of principle for key technology
	Construction of evaluation criterion and indicator system
	Applied research for assessment of the project technology maturity
	Evaluation of the pilot and extension service for the project technology maturity
Project investment management	Assessment for investment feasibility
	Investment examine and approve
	Investment supervise and check up
	The assessment of investment effect
Achievement and intellectual property management	The policy of the intellectual property management and the research of the institutional system
	The key cycle of management about the assessment of the intellectual property problems
	The research on achievement and intellectual property management
Risk management	Risk management of project technology
	Risk management of project cost
	Risk management of project process
	Risk management of project quality
Technical support management	The analysis of technical support management factors
	The analysis of technical support management flow
	The design of auxiliary methods for technical support
Security and secrecy management	Secret-level setting
	The construction of confidentiality rules and regulations
	The design of secrecy policy
	The control of security process management
	The design of scheme for security technology
Decision support management	Auxiliary decision support analysis for the project
	Decision correlation analysis for the project
	The design of the method for the implementation of the decision support

The full meaning in Figure 2: PPAPM represents project planning and project management, PM represents project management, PQCAM represents project quality control and management, PTMM represents project technology maturity management, PIM represents project investment management, AAIPM represents achievement and intellectual property management, RM represents risk management, TSM represents technical support management, SASM represents security and secrecy management, DSM represents decision support management.

C. The quantification of national key science and technology program of China evaluation index system

The implementation of the national key science and technology programs of China is effected by many factors; it may lead some unpredictable problems if these factors are analyzed only by experts' experiences. Therefore, evaluation factor system is assessed by the methods of Analytic Hierarchy Process [7] and Entropy Method [8], while the management maturity level of the national key science and technology programs of China is assessed by utilizing Fuzzy Comprehensive Evaluation [9]. The design of the evaluation system is qualitative analysis, by utilizing the methods such as Analytic Hierarchy Process, it combines qualitative analysis with quantities analysis; it compensates the defect of quantities analysis.

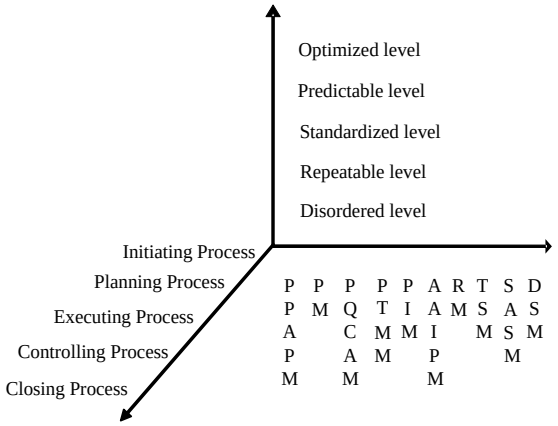


Figure 2. The three dimension structure of national key science and technology program of China

The quantification of the model of the national key science and technology programs of China is mainly assessed by Analytic Hierarchy Process and Entropy Method. The experts which have project management experience score the evaluation index system for each index, then the score of each evaluation index is arranged, the evaluation index is selected by the score of each index, and the final evaluation index system is formed. Then, the participants of the project score the weights of each management factors and a questionnaire about maturity level, the participants of the project are those people who have participated some projects; for example, if we want to assess the management maturity level of China High resolution Earth Observation System, we choose those people who have participated the project of China High resolution Earth Observation System to score the weights of each management factors and a questionnaire about maturity level, then, the model of national key science and technology program of China is built through calculating the weights of each factors, the score and the weight present the importance of each evaluation index, the model can be used to evaluate the

 ISBN: 978-1-61208-317-9

maturity level of the national key science and technology programs of China.

III. CASE STUDY

A. The profile of China High resolution Earth Observation System

China High resolution Earth Observation System (CHEOS), one of the 16 national key science and technology programs of China is selected as a case study; the aim of this program is mainly to develop the satellite, airplane and stratospheric airship's high resolution earth observation. With the use of medium and low resolution earth observation, a high resolution earth observation with all-weather, all-time capabilities are formed.

The implementation of the CHEOS is aimed at enhancing the ability of earth observation, mastering the initiative of information and resource and grasping the situation of economy, resource, and environment, CHEOS is an essential system to our nation, it has a significant meaning for protecting our nation and enhancing national strength.

B. The confirm of evaluation index and evaluation level

(1) The confirm of evaluation index system

There are 10 first-class targets and 41 second-class targets in the maturity model of CHEOS, the details of each level: $U=\{U_1,U_2,U_3,U_4,U_5,U_6,U_7,U_8,U_9,U_{10}\}$={project planning and project management, project management, project quality control and management, project technology maturity management, project investment management, achievement and intellectual property management, risk management, technical support management, security and secrecy management, decision support management}.

(2) The confirm of maturity level

The maturity level of CHEOS: $V = \{V_1,V_2,V_3,V_4,V_5\}$ = {Disordered level, Repeatable level, Standardized level, Predictable level, Optimized level}.

C. The confirm of the weights of evaluation index

1) The confirm of initial weights via Analytic Hierarchy Process

(1) The hierarchical structure of evaluation index system

The evaluation index system is divided into 3 levels: target layer, criterion layer and factor layer.

(2) Comparison and judgment matrix

The judgment matrix is designed through questionnaire and expert's graded approach, the value of each matrix reflects the degree of importance of each index.

The model uses a_{ij} to represent the relative importance between i and j, $U=(a_{ij})_{m\times n}$ represents comparison and judgment matrix.

$$U=\begin{bmatrix} U_1/U_1 & \cdots & U_1/U_{10} \\ U_2/U_1 & \cdots & U_2/U_{10} \\ \vdots & \vdots & \vdots \\ U_{10}/U_1 & \cdots & U_{10}/U_{10} \end{bmatrix}=\left[a_{ij}\right]_{n\times n} \quad (1)$$

(3) Calculation of the weight

The model uses the formula of $UW=\lambda_{max}W$ to calculate the eigenvalues, λ_{max} is the eigenvalue of U, W is the weight of evaluate factors.

2) Update weights via Entropy method

(1) The fundamental of Entropy method

In information theory, information entropy is a concept to measure the amount of information. If a system is very orderly, the value of information entropy is small, and the reverse situation is also true. So, information entropy is a way to measure a system's degree of order [6], the expression is:

$$H(x)=-\sum_{i=1}^{m} p(x_i)\ln p(x_i) \quad (2)$$

x_i represents the state of i, $p(x_i)$ represents the probability of i.

For example, there are m experts score n evaluation factors, the judgment matrix is $X=(x_{ij})_{m\times n}$, j is a factor, if the value between each factor is large after the expert's assessment, the weight of the factor should be high, conversely, the weight of the factor should be low. So, the weight of each factor can be adjusted through Entropy Method.

(2) Weight adjustment

Step 1: calculate the weight of x_{ij}

$$p(x_{ij})=x_{ij}/\sum_{i=1}^{m} x_{ij} \quad (3)$$

Step 2: calculate the entropy of e_j

$$e_j=-k\sum_{i=1}^{m} p(x_{ij})\ln p(x_{ij}) \quad (4)$$

Step 3: calculate the otherness factor of g_j

To the factor of j, if the otherness of x_{ij} is small, and the entropy of e_j is big; when the entropy of e_j is equal to e_{max}, the factor of j makes no difference; when the value of each factor

 ISBN: 978-1-61208-317-9

is big and the entropy of e_j is small, the factor contributes more to the evaluation system.

Define $G=(g_1,g_2,\ldots,g_n)$ as otherness coefficient vector, among $g_j=1-e_j$, if the otherness factor of g_j is big, this factor plays an important role in the model.

Step 4: weight adjustment

After the factor of b_j is calculated through Analytic Hierarchy Process, the factor is adjusted by utilizing the otherness coefficient of g_j.

$$a_j = b_j \times g_j, j = 1,2,3,...,n \quad (5)$$

The weigh of the factor is obtained after the normalization of a_j

$$w_j = a_j / \sum_{j=1}^{n} a_j, j = 1,2,3,...,n \quad (6)$$

D. The assessment of the maturity level of CHEOS

(1) There are 5 maturity levels in CHEOS: $V=\{V_1,V_2,V_3,V_4,V_5\}$={Disordered level, Repeatable level, Standardized level, Predictable level, Optimized level}

The value of each maturity level is: Disordered level = 1, Repeatable level = 2, Standardized level = 3, Predictable level = 4, Optimized level = 5.

The weight vector is: $V=(1,2,3,4,5)^T$

(2) The weight of each criteria layer

The weight of criteria layer and factor layer are presented by vector through Analytic Hierarchy Process and Entropy Method: Define $W = (W_1,W_2,W_3,\ldots,W_{10})$ as the weight of criteria layer, define $w_1 = (w_{11},w_{12},w_{13},w_{14},w_{15},w_{16}),\ldots,w_{10} = (w_{101},w_{102},w_{103})$ as the weight of each factor layer.

(3) The confirmation of judgment matrix

For example, to the technical support management of criteria layer, the judgment matrix is:

$$R_8 = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{15} \\ r_{21} & r_{22} & \cdots & r_{25} \\ r_{31} & r_{32} & \cdots & r_{35} \end{bmatrix}$$

$r_{ij}(j=1,2,3,4,5)$ represents the probability of r_i which scores j:

$$r_{ij} = \frac{\text{number of the questionnaire which i scores j}}{\text{number of the questionnaire}} \quad (7)$$

The judgment vector of B_8 is calculated by the weight of w_8 and the judgment matrix of R_8:

$$B_8 = w_8 \times R_8 = (w_{81}, w_{82}, w_{83}) \times \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{15} \\ r_{21} & r_{22} & \cdots & r_{25} \\ r_{31} & r_{32} & \cdots & r_{35} \end{bmatrix} = (b_{81}, b_{82}, \ldots, b_{85}) \quad (8)$$

E. The method to analysis and improvement the management factors

S_i is a vector of criteria layer *i (i=1,2,3,…,10)*:

$$S_i = R_i \times V^T \quad (9)$$

The coordinate system is divided into 4 areas in order to improve the indexes of CHEOS, X axis represents the score of each index, and Y axis represents the weight of each index, then the indexes are labeled in the coordinate system, the management index is improved via analyzing the points in the coordinate system. The coordinate system is as follows in Figure 3:

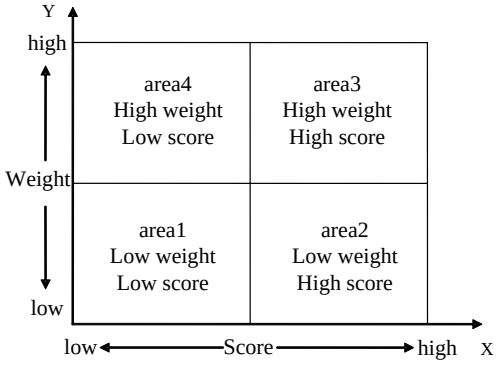


Figure 3. The analysis and improvement of the management factors

Area 1: Low weight, low score

The management activities in this area have a low weight and low score, low score illustrates the project manager pays little attention to these management activities. Low weight illustrates these management activities have poor importance in the process of management, according to the coordinate system that can conclude these management factors are the weaknesses of the CHEOS, compare with the activities in area 4, those activities also has a low score, but the weight of these activities in area 1 is lower than area 4, so the management of CHEOS should pay more attention to area 4. And the activities in area 1 should be improved correspondingly.

Area 2: Low weight, high score

The activities where belong to this area have a low weight, in compare with a high score, manager pays enough attention to these kind of activities, but these activities are not so important compare with the activities in area 4 because of its

 ISBN: 978-1-61208-317-9

low weight. So, more attention should be paid in other area other than the activities in area 2.

Area 3: High weight, high score

The activities in this area both have a high weight and score. These activities play an important role in project management, and manager pays enough attention to these activities. These activities don not need to be improved so frequently, but these activities have a significant meaning in the national key science and technology programs of China, so the management should focus on these activities continuously.

Area 4: High weight, low score

The activities in area 4 have a low score, but the weight of these activities are high, so the activities in this area have an important function in the process of CHEOS, but the management of the CHEOS pays not enough attention to these activities, so the management ability in these aspects are poor. So, these activities are the objects which the project mainly improves. The improvement of these activities can promote the maturity level of the CHEOS and urge the implementation of the project successfully.

IV. CONCLUSION AND FUTURE WORK

This paper studied the modeling of the project management maturity for the national key science and technology programs of China by updating the current management maturity models. The applicability of the proposed management maturity model with consideration of the features of the programs is demonstrated through a real case study, which proves the compatibility of the proposed model with the present conditions of China.

The paper aims to introduce the maturity model into the management of the national key science and technology programs of China. According to the process and the result of research, the evaluation index system is to be improved through combining more management features of the national key science and technology programs. Although the maturity model is successfully introduced into the management of the programs, the method of assessment is still simple, mostly analytic hierarchy process or fuzzy comprehensive evaluation; so, that further work is planned to make improvement.

In conclusion, the management of the national key science and technology programs of China is a complex systems engineering issue. The introduction of the maturity model is just an explorative trial. Therefore, more theoretical and practical research on the application of the maturity model for the management of the national key science and technology programs of China is still needed.

ACKNOWLEDGEMENTS

The authors would like to acknowledge the support of National Key Science and Technology Program of China High Resolution Earth Observation System (grant No. B0601/01-1010-2010).

REFERENCES

[1] R. Y. Cheng and Y. Wang, “The main features of the national key science and technology programs of American,” Journal of Research on Technological Management, 2008, vol. 6, pp. 38-40.

[2] J. X. Wu and C. Dong, “The features and enlightment of the national key science and technology programs of foreign country,” Journal of Research on Technological Mangement, 2011, vol. 9, pp. 29-32.

[3] B. M. Hu, T. Wang, and Z. B. Li, “The research on the features of the national key science and technology programs,” China technological forum, 2007, vol. 9, pp. 81-85.

[4] Y. L. Zhang and J. L. Jin, “The application analysis for the management of the maturity model,” The development of technologies in enterprise, 2012, vol. 20, pp. 69-74.

[5] H. Kerzner, “Strategic planning for a project management,” Project Management Journal, 2003, vol. 2, pp. 54-60.

[6] D. P. Wang and X. Wang, “Research on weights about green suppliers in iron and steel enterprise through AHP and Entropy Method,” Soft Science, 2010, vol. 8, pp. 117-122.

[7] J. Xu and N. Liu, “The basic idea and the practical application of the Analytic Hierarchy Process,” Intelligence Explore, 2008, vol. 12, pp. 76-80.

[8] Z. Y. Ma, “Comprehensive evaluation model based on improved entropy method,” China Science and Technology Information, 2009, vol. 17, pp. 215-220.

[9] M. Yao, “An improved fuzzy comprehensive evaluation method,” China Soft Science, 1990, vol. 1, pp. 1-6.

 ISBN: 978-1-61208-317-9

Optical Multicast Protocol for LEO Satellite Networks

Maha Sliti, Walid Abdallah, Noureddine Boudriga
Communication Networks and Security Research Lab, University of Carthage
Ariana, Tunisia
e-mail: maha.sliti@supcom.tn, wabdallah_tn@yahoo.fr, nab@supcom.rnu.tn

***Abstract*—Satellite networks provide global coverage and support a wide range of services. Since Low Earth Orbit (LEO) satellites provide short round-trip delays, they are becoming increasingly important for real-time applications, such as voice, teleconferencing and video traffic. Many applications require a mechanism to deliver information to multiple recipients. In this context, we propose two all optical multicasting approaches based on codewords designed for LEO satellites: the first approach is based on the shortest paths and the second approach is based on the concept of virtual multicast trees. The proposed multicast approaches consider free space optical communication links between satellites, which provides high data rate for real time applications.**

***Index Terms*—LEO satellite, all optical multicast, optical codewords, tunable optical decoder.**

I. INTRODUCTION

Satellite networks provide global coverage and support a wide range of services. Since Low Earth Orbit (LEO) satellites provide short round-trip delays, they are becoming increasingly important for real-time applications, such as voice, teleconferencing and video traffic. Many applications require a mechanism to deliver information to multiple recipients. In the literature, several multicast routing protocols were proposed for satellite networks [1]–[4]. However, the majority of the proposed approaches are not developed primarily for LEO satellite constellations. In fact, they are mainly proposed in the context of IP and mobile networks and then adapted to support multicasting in LEO networks. Furthermore, the proposed approaches have not consider free space optical communication links between satellites, which are ideal for high data rate real-time applications, such as voice and video traffic.

In this paper, we propose two all optical multicasting approaches based on codewords designed for LEO satellites: the first approach is based on the shortest paths and the second approach is based on the concept of virtual multicast trees. Furthermore, we study the mobility management aspect for the two proposed approaches. The key contributions of the proposed multicast approaches with respect to the previously published research are:

1) The proposed multicast process is performed at the optical layer based on an optical switching concept, which allows the multicast packets to be processed at very high bit rates (gigabits per second) without conversion to the electronic domain. During the optical multicast process, the received packets are delayed in an optical buffer proposed in [5], in order to provide a tunable delay for real time traffic.
2) The optical switching concept is based on the optical codewords, which are represented by a sequence of pulses. Indeed, a codeword is assigned to each satellite in the network and serves as an optical identifier of the satellite. Based on the received codeword and the structure we build in, the traffic will be multicasted to one or several directions allowing to reach the destination satellites.
3) The proposed multicast approach is scalable since its performance is not affected by the multicast group size and the member combination in the multicast group. Therefore, the optical multicast module implemented in each intermediate satellite is at most composed of three tunable decoders.

The rest of the paper is organized as follows. Section 2 presents the proposed multicast approaches for the Leo constellation networks in the literature. The use of optical codewords for the multicast in LEO networks is described in Section 3. The all optical multicast approach based on the shortest paths is described in Section 4. The all optical multicast approach based on the virtual multicast trees is explained in Section 5. The mobility management in the two all optical multicasting approaches is discussed in Section 6. Simulations and experimental results are given in Section 7. Finally, Section 8 concludes the paper.

II. MULTICAST IN LEO CONSTELLATION NETWORKS

Satellite-based optical communication systems have become a most promising technology for high-speed inter-satellite and satellite-to-ground communication links due to their various advantages. A free space optical communication system includes optical transmitter and receiver satellites. The transfer of information between the two satellites is performed with optical rather than microwave radiation, which allows high data rate transfer and transparency to Radio Frequency (RF) interference. The advantages of an optical communication system compared with a microwave communication system in free space are: 1) high data rate, 2) less transmitter power consumption, 3) terminal design with reduced size and weight, and 4) transparency to RF interference.

 ISBN: 978-1-61208-317-9

In this paper, we are particularly interested in LEO satellite mobile systems. LEO satellites are orbiting at low earth orbits with an altitude generally between 500 km and 2000 km. Compared to communication satellites in geostationary orbit, the communication links to LEO satellites are characterized by lower propagation delay and lower link attenuation because of the shorter distance, resulting in the need for reduced transmission power.

In the literature, several multicast routing protocols were proposed for satellite networks. In [1], a new core-based shared tree algorithm, viz Core-Cluster combination-based Shared Tree (CCST) algorithm and the weighted version (i.e., w-CCST algorithm) are proposed in order to resolve the channel resources waste problem in typical source-based multicast routing algorithms in LEO satellite IP networks. The Rectilinear Steiner Trees (RST) [2] algorithm uses one of integer linear programming method for solving the minimum RST tree, so that the bandwidth of multicast routing using the least suitable non-real-time multicast applications, but more complicated calculation process. In [3], a fast iterative distributed multicast routing algorithm was developed based on the inherent characteristics of satellite networks, that consider distributed computing model and significantly reducing the algorithm computational complexity. In [4], a QoS-Guaranteed Secure Multicast Routing Protocol (QGSMRP) is proposed for satellite IP networks using the logical location concept to isolate the mobility of LEO and HEO satellites. a novel triple-layered satellite network architecture including Geostationary Earth Orbit (GEO), Highly Elliptical Orbit (HEO), and Low Earth Orbit (LEO) satellite layers is introduced. However, the majority of the proposed approaches are not developed primarily for LEO satellite constellations. In fact, they are mainly proposed in the context of IP and mobile networks and then adapted to support multicasting in LEO networks. Furthermore, the proposed approaches have not consider free space optical communication links between satellites, which are ideal for high data rate real-time applications, such as voice and video traffic.

III. CODEWORD BASED MULTICAST

An optical codeword is composed of a sequence of pulses that represent the positions of the information bit "1". The transmitter does not produce any optical pulse when the information bit "0" is transmitted.

An optical codeword is assigned to each satellite in the LEO constellation network. Therefore, each satellite is uniquely identified by its assigned codeword which is similar to an address in our case. The total number of associated codewords is equal to $N * M$ with N is the number of orbits (planes) in the constellation, and M is the number of satellites in each orbit. We notice that the number of codewords used for a LEO constellation network (e.g., Global Positioning System (GPS)) is very reduced compared to current terrestrial networks (e.g., 24 codewords are required in the GPS network). We suppose that the codewords are generated, assigned and managed by a central entity implemented in a terrestrial station which is the ground station in our case.

To allow the application of our physical routing approach and discriminating data transmitted to different satellites, the optical codes should be orthogonal. In our case a codeword can be represented as an integer vector where each element identifies the position of a pulse (bit "1"). Thus, we adopted a generation process based on lattice point theory to satisfy this requirement. The code generation process is realized as follows: Consider a $m \times k$ integer lattice $L = Z_m \times Z_k$, which elements are labeled by points from the set $V = 1, 2, ..., mk$. Using a simple linear mapping function defined as:

$$l : L \rightarrow V, l(x, y) = mx + y + 1 \quad (1)$$

where $0 =< x <= k - 1$ and $0 =< y <= m - 1$.

The codewords are represented as lines connecting points of the rectangular integer lattice. The subset of points are referred to as lines, and the code spaces are defined as the sets of lines of different slopes. A line with a slope s, where $0 \leq s \leq m-1$, starting at the point (i, j), contains the following set of points:

$$(i; j + (s.imod(m)) : 0 \leq i \leq k - 1; 0 \leq j \leq m - 1 \quad (2)$$

Thus, for every slope s, the m codewords that can be obtained are $\nu = \nu_1, ..., \nu_m$, where $\nu_j = mi + j + (s.imod(m)) + 1; 0 \leq i \leq k - 1$ for every $0 \leq j \leq m - 1$. It can be shown in [6] that this method generates codewords with length equal to mk and weight k. The weight witch represents the number of bits "1" in the codeword. This is a very important parameter because it identifies the maximum value of the autocorrelation between codes and hence reflects the capability of the routing approach to separate traffics intended for specific destinations. It is noteworthy that, for every slope, the codewords obtained through the aforementioned process are orthogonal. According to this reasoning, every codeword ν_j can be represented as an integer vector $\nu_j = \nu_j^0, \nu_j^1 ..., \nu_j^{k-1}$ where $j \in 1, 2, ..., mk$ and ν_j^i is the i^{th} bit in the optical code ν_j.

Since LEO satellites provide short round-trip delays, they are becoming increasingly important for real-time applications, such as voice and video traffic. Many applications require a mechanism to deliver information to multiple recipients, as illustrated by Figure 1.

Since each group of destination users has a geographic location, each group is covered by a different distribution satellite that we call also a destination satellite. At $t0$, a source user is attached to a source satellite and a group of destination users is served by a destination satellite via a wireless link. On the contrary, the links between satellites, which are free space optical communication links. During the communication period, the group of destination satellites are changed due to the mobility of the source and destination satellites. Thus, the management of the multicast process between the source and the destination satellites must be done during communication periods that may last for hours.

 ISBN: 978-1-61208-317-9

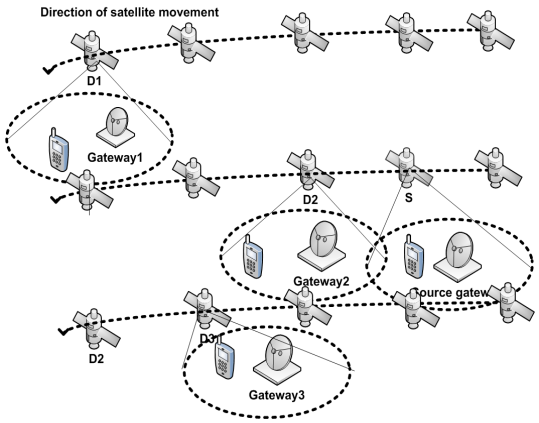


Figure 1. Multicasting in LEO satellite constellation.

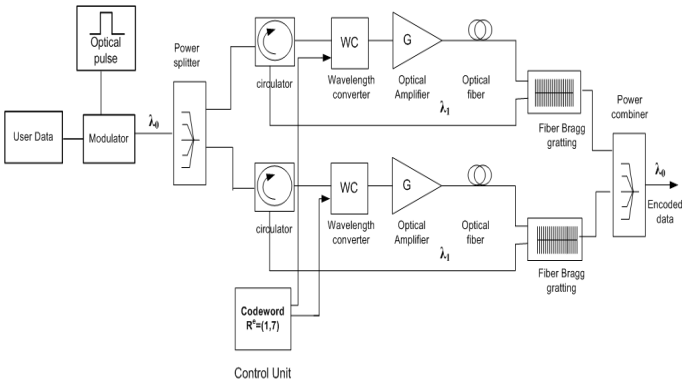


Figure 2. Tunable optical decoder architecture.

In this work, we propose an optical multicasting process based on codewords. Also, we provide a function so that each satellite can optically switch a multicasting traffic based on codewords properly structured. Thus, each satellite implements an optical switching module based mainly on two optical operations, which are the optical codeword matching and the deviation of the traffic to the adequate direction. The received codeword is split to a set of decoders; if a codeword is matched by a decoder, an optical switching gate will be activated by a pulse to forward the multicast traffic to the direction that allows reaching the destination satellite.

The physical implementation of the reconfigurable encoding is achieved by delaying pulses in a set of Optical Delay Line (ODL) loops. The positions of the pulses in the codeword define the number of rounds that must be performed by every pulse in each ODL-loop [7]. The optical sequence of code is then created by combining different delayed pulses at the output of ODLs.

A codeword decoder, which is the main component of the proposed optical multicasting module, allows the matching of a codeword optically without the O/E/O conversion. The design of a tunable decoder was proposed in [7] in order to implement an optical filtering technique based on codewords. In each loop, we inject the optical pulse, which will be confined in the fiber until it performs the number of rounds needed to place it in the positions of a '1' bit in the optical code. It is obvious that the number of ones in the code sequence determines the number of optical loops needed. Consequently, we could easily control and modify the codeword by controlling the number of rounds performed by the optical signal in every optical loop. Unlike classical techniques, such as fixed length Optical Delay Lines and Fiber Bragg Gratings, in which modifying optical codes needs a physical intervention, our approach allows soft based controlling of the optical codewords by considering a control unit, even if it uses Fiber Bragg Gratings. A codeword is considered as an integer vector of k elements that presents the positions of ones in the codeword. Thus, an optical codeword is composed of optical pulses transmitted in specific bit intervals that correspond to positions of ones in the code sequence. This can be achieved by considering an on-off keying technique and modulating the label by a Gaussian optical pulse signal and a Mach-Zehender modulator. The tunable decoder, illustrated by Figure 2, generates an optical pulse on the receipt of a valid codeword that matches the configured code-word. The decoding operation is performed by delaying optical pulses that compose the received codeword until they superpose in the last bit interval.

IV. ALL OPTICAL MULTICAST BASED ON THE SHORTEST PATHS

In this section, we present an optical multicast approach based on the shortest paths. In this approach, each intermediate satellite considers three codewords indicating its direct neighbors: two in the same orbit, and one in the following orbit. A shortest path is defined in term of number of hops and is established by favoring the inter-orbit on the intra-orbit hop. In order to establish the shortest paths to a list of destination satellites, the route discovery process is initiated by the source satellite. The source satellite duplicates a Route REQuest message (RREQ) in order to send it to d destination satellites. The considered RREQ message format is composed of: the message identifier, the codewords associated to destination satellites, the satellite source address, the satellite destination address, and the communication time between the source user and the destination users.

At the reception of a RREQ message, the intermediate satellite adds its codeword to the Multicast list address and sends it to the nearest neighbor based on the Destination address in the RREQ message. A destination satellite that receives the RREQ message, sends a Route RESponse (RREP) message, which contains the shortest path to the source satellite. A path is formed by a list of codewords that denote the intermediate nodes in the shortest path.

After the route discovery process, a source satellite sends the multicast traffic to the d destination satellites. Each multicast packet is duplicated on the shortest paths established to destination satellites. The paths are composed of a list of codewords that characterize the list of intermediate nodes on

 ISBN: 978-1-61208-317-9

the shortest path. Thus, a header that contains the path to a destination satellite is associated to each packet.

The design of the multicast module implemented in each satellite is illustrated by Figure 3. Therefore, an intermediate node that receives a multicast packet examines the received header optically by considering the following steps:

- the first codeword in the header, which indicates the current satellite, is extracted from the received list of codewords that compose the header and dropped;
- the packet and the new header are delayed in a Virtual Optical Memory (VOM) based on optical delay lines, that we have developed in [5];
- during the buffering delay, the second codeword in the received header is extracted from the received list of codewords and split to three tunable decoders. Each decoder allows to match a codeword that characterize one neighbor of the current satellite;
- If the second codeword matches a configured codeword, then the delayed packet and its corresponding header will be sent to the adequate next neighbor that allows to reach the destination satellite.

V. All optical multicast based on multicast trees

In this section, we present an optical multicast approach based on the concept of virtual multicast trees. The virtual multicast paradigm is used in order to underline that the multicast tree is not physically established but only built on the codewords structure used to switch traffic contrary to the first proposed approach that requires a route discovery process preceding the routing of a multicast traffic. In this approach, the virtual tree establishment consists on the management of codewords structure. Thus, an optical codeword is structured in two parts as follows: the first part identifies the orbit and the direction (left or right) in this orbit, and the second part identifies uniquely a satellite in the LEO constellation network. Therefore, the destination satellite can be either on the left or on the right of an intermediate satellite or it can be localized in another orbit. The source satellite sends a multicast packet composed of the payload and a list of codewords that corresponds to the list of destination satellites. At the reception of a multicast traffic, an intermediate satellite can forward the traffic to the right or left if the destination satellite is in its orbit or it switches the traffic to the following orbit.

In the following example, we consider the virtual tree structure illustrated by Figure 4. Therefore, from a source satellite, the header to be sent, which is composed of a set of codewords, has the following structure: $CdOrbit_r^2$ CdD_1| $CdOrbit_l^2$ CdD_2| $CdOrbit_l^2$ CdD_3 | $CdOrbit_l^3$ CdD_4 | $CdOrbit_r^3$ CdD_5, where $CdOrbit_r^i$ is the codeword that identifies the right direction in the $Orbit^i$, $CdOrbit_l^i$ is the codeword that identifies the left direction in the $Orbit^i$, and CdD_j is the codeword that identifies the destination satellite D_j in the LEO constellation network.

Each intermediate satellite integrates a multicast module, which design is described by Figure 5. The optical multicast process is performed as follows:

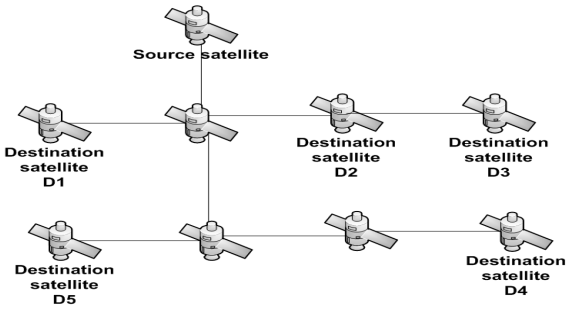


Figure 4. Virtual tree structure.

- The received payload is extracted and delayed in the optical buffer called Virtual Optical Memory (VOM), proposed in [5];
- The list of received codewords are extracted and separately treated by the multicast module. If the intermediate satellite is not a destination satellite, it split the received codewords to three tunable decoders. Each decoder allows to match a received codeword in the header. The matching operation is performed in serial by treating each time one codeword until treating all the codewords that compose the received header. And, the matching is performed only on the first part of the codeword that indicates the switching direction: left or right if the destination satellite is in the same orbit as the intermediate satellite, or down if the destination satellite is in the following orbits. If the intermediate satellite is a destination satellite, based on the received codewords, it performs the matching operation based on the other received codewords in the header and does not consider its codeword. Indeed, a satellite needs to perform a matching operation on the second parts of the received codewords that uniquely identify a satellite in the LEO constellation network before performing the optical switching operation based on the matching of the first parts of codewords that identify the forwarding direction.
- If a codeword is matched then an SOA gate (Semiconductor Optical Amplifier gate) is activated in order to pass the delayed payload to the new header generation module. This module generates a new header for the payload by considering only the codewords of the satellites that are reachable by the forwarding output port. Indeed, the group of destination satellites is divided in each intermediate satellite based on the fact that the traffic that is forwarded to a direction (left, right, down) must only contain in its header the codewords corresponding to the destinations located in this direction.

VI. Multicast coping with long period services

Satellite movement results in challenging mobility management problems in LEO satellite networks. Due to the movements of the satellites and according to the movements

 ISBN: 978-1-61208-317-9

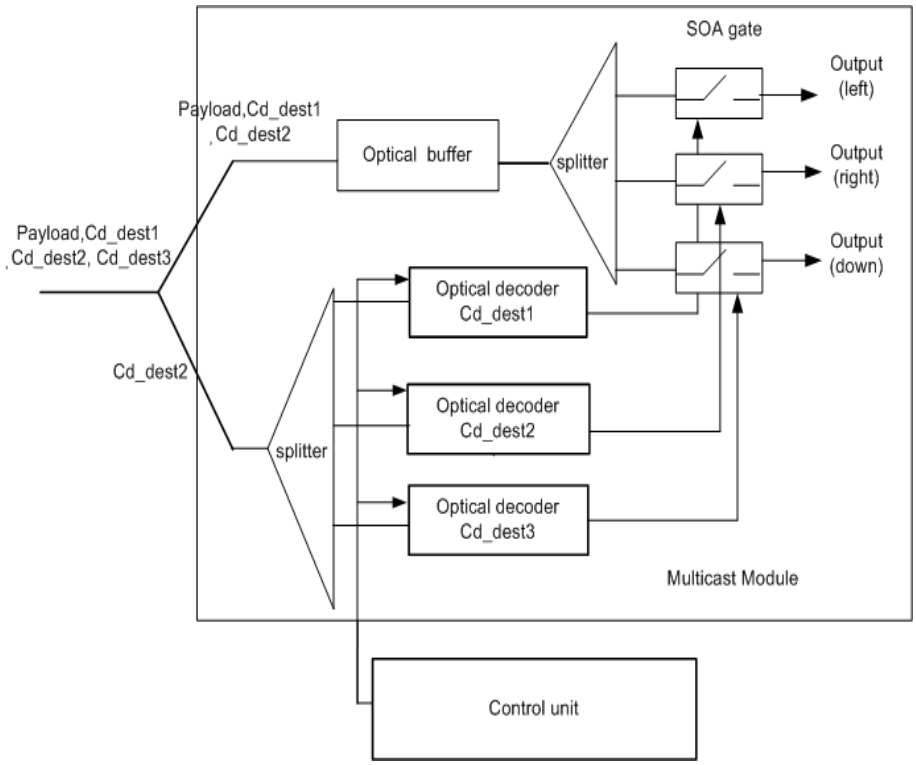


Figure 3. All optical multicast based on the shortest paths.

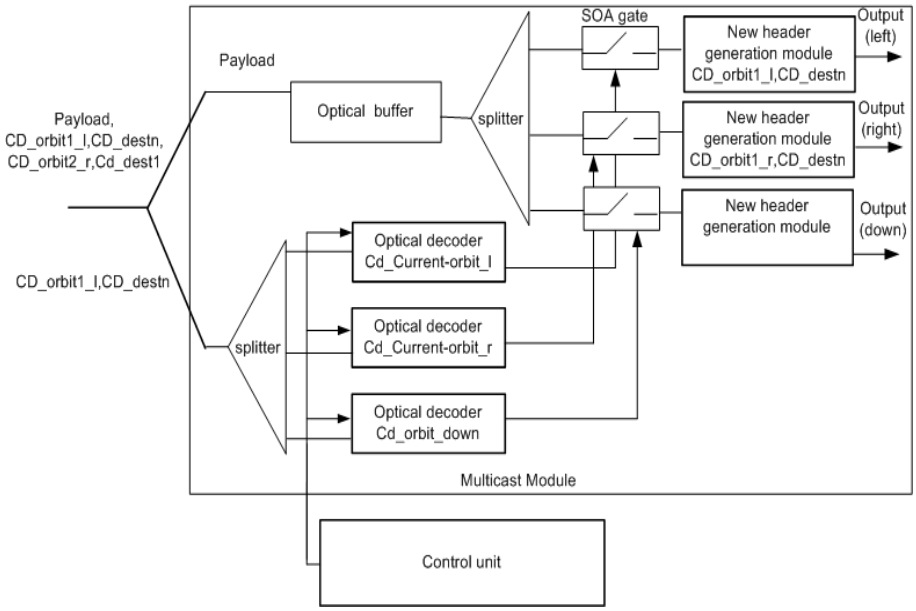


Figure 5. All optical multicast based on the Virtual multicast tree.

of their coverage area and footprints, a group of destination users are served by several groups of satellites during a communication period that may lasts for hours. In this section, we present the mobility management for the two proposed optical multicasting approaches. For the two approaches, three types of handovers can occur during a multicast session period T. The first type of handover occurs when the source satellite moves and will not serve the users source of the traffic anymore. The second type of handover occurs when a destination satellite will not serve any destination user therefore it will be deleted from the group of the destination satellites. The third type of handover occurs when a destination satellite is added

to the group of the destination users in order to serve one or several destination users.

A. Shortest paths management

The optical multicasting based on the Shortest paths requires the initiation of a novel Route request discovery process to the destination satellites when a source satellite is not in the coverage of source users anymore. The ground station must send the current list of codewords, that indicate the current list of destination satellites, to the new source satellite. In the case where a destination satellite will not serve any destination user, it will be deleted from the group of destination satellites. This type of mobility does not require any treatment in the intermediate satellites. In fact, the multicast traffic to be switched by the intermediate satellites will no longer be switched in the direction of the deleted satellite since it is not in the coverage of the destination users. An updated list of codewords, that corresponds to the new list of destination satellites, will be sent to the source satellite from the ground station. In the case where a destination satellite is added to the group of destination users in order to serve one or several destination users, the shortest path from the source to the new destination satellite must be established. An updated list of codewords will be sent to the source satellite from the ground station.

B. Virtual tree management

When a source satellite is not in the coverage of source users anymore, the optical multicasting based on the virtual tree concept does not require any route reestablishment process initiation to discover the routes to the destination satellites as it is the case in the optical multicasting based on the Shortest paths. Indeed, the multicast tree will be implicitly established by sending a multicast traffic that requires at each intermediate satellite to be optically switched. The ground station must send the current list of codewords, that indicate the current list of destination satellites, to the new source satellite. In the case where a destination satellite will not serve any destination user, it will be deleted from the group of destination satellites. The ground station must send the current list of codewords to the new source satellite. And, the source replaces the codeword corresponding to the deleted destination satellite by the codeword associated to the new destination satellite and adds to the codeword, a special codeword that identifies the destination orbit and the direction (left or right) in the orbit. The new codeword will be added to the header of the multicast traffic. In the case where a destination satellite is added to the group of the destination users in order to serve one or several destination users, its corresponding codeword will be added to the header of the multicast traffic.

C. Comparison

First, the two proposed optical multicasting approaches are compared in term of the number of used segments. The segments used by the approach based on virtual trees are those in the tree axis and in the left and right of the axis. And, the segments used by the approach based on the shortest paths are the total segments that compose the shortest paths to the destination satellites. Thus, the number of segments used in the first approach avoids the segments redundancy and consequently it is smaller than the number of segments used in the second approach.

Second, the virtual tree construction favors the inter-orbit over the intra-orbit hop, which allows to have only one possible path to the destination satellite. This method is also considered in order to establish the shortest paths in the second approach. Indeed, the association of all established shortest paths gives the constructed virtual tree in the first approach.

Third, in the approach based on the shortest paths, when a satellite leaf handover occurs in the direction of satellites movement, a segment is removed of the path and the codeword of the removed destination satellite is removed of the list of destination satellites. In the contrary case, a segment will be added to the path and the codeword of the new destination satellite is added to the list of destination satellites. Thus, the handover in this approach consists on the increase or the narrowing of paths. When we consider the approach based on virtual trees, the destination satellites can be either a leaf or an intermediate satellite, which minimizes the increase or the narrowing of paths due to handovers compared to the first approach.

VII. Simulations and experimental results

In this section, we assess the performance of the two proposed optical multicasting approaches in terms of traffic load (in erlang) and number of decoding functions. As simulation environment, we have considered Matlab tool. For simulation purpose, we consider: the Global Positioning System (GPS) constellation topology, which forms a LEO network (6 orbits, each orbit has 4 satellites), a traffic load matrix. The traffic load matrix is a matrix that considers a traffic load only between each satellite and its four neighbors (up and down, left and right). For each simulation, a source satellite number and destination satellite numbers are randomly generated.

In order to compare the efficiency of the two multicasting approaches in terms of the traffic load criteria, we calculate the traffic load for each approach by considering the maximal path length/multicast tree depth and the mean on all shortest paths/segments of the tree. As illustrated by Figure 6, we can notice the similarity between the traffic load curves computed on the maximal path length and the depth of the multicast tree and considered for several multicast group sizes (4,8,12,16,20,24). This similarity can be justified by the fact that the maximum path length corresponds to the depth of the multicast tree. We notice also that the mean traffic load on the total segments of the multicast tree is the half for a multicast group size equals to 24 satellites, which is optimized compared to the mean traffic load on the total path lengths. Therefore, we can deduce that the all optical multicasting approach based on the virtual multicast tree has the advantage to eliminate the redundant segments in the shortest paths established in the all optical multicasting approach based on the shortest paths.

 ISBN: 978-1-61208-317-9

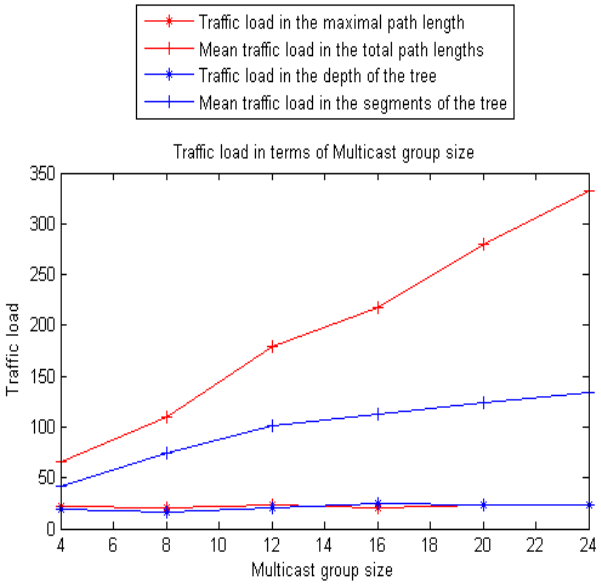


Figure 6. Traffic load in terms of multicast group size.

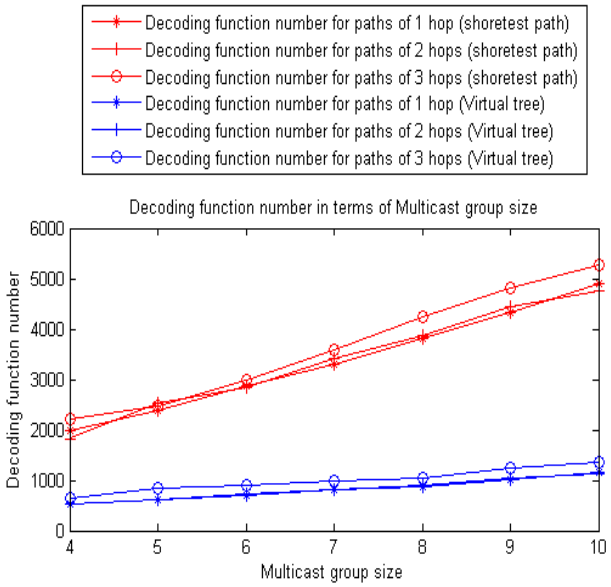


Figure 7. Decoding function number in terms of multicast group size.

The efficiency of the two multicasting approaches in terms of the number of used decoding functions is also assessed. We have considered a communication period of one hour between a source and destination users. During this period and due to the mobility of satellites, four source satellites and four multicast groups are considered. We calculate the number of used decoding functions for: 1) different multicast group sizes (4,5,6,7,8,9,10), 2) three path lengths less or equal to 1,2,3 hops. As illustrated by Figure 7, we notice that the number of used decoding functions for optical multicasting approach based on the shortest paths is five times greater than the number of used decoding functions for optical multicasting approach based on the virtual multicast tree for a multicast group size equals to 10 satellites. This can be justified by the fact that the virtual multicast tree eliminates segments redundancy observed in the shortest paths approach. Consequently, the optical multicasting approach based on the virtual tree decreases the traffic load observed in the network segments and uses less decoding functions for the optical multicasting. Thus, this approach is more efficient than the approach based on the shortest paths.

VIII. Conclusion

In this work, we have proposed two all optical multicasting approaches based on codewords designed for LEO satellites: the first approach is based on the shortest paths and the second approach is based on the concept of virtual multicast trees. For the two approaches, we identify each satellite in the LEO constellation network by an optical codeword. The proposed multicast process is performed at the optical layer based on an optical switching concept, which allows the multicast packets to be processed at very hight bit rates by considering optical codewords. Indeed, a codeword is assigned to each satellite in the network and serves as an optical identifier of the satellite. Based on the received codeword, the traffic will be multicast to one or several directions allowing to reach the destination satellites. Furthermore, the proposed multicast approach is scalable since its performance are not affected by the multicast group size and the member combination in the multicast group. Therefore, the optical multicast module implemented in each intermediate satellite is at most composed of three tunable decoders.

References

[1] L. Chen, J. Zhang, and K. Liu, "Core-Based Shared Tree Multicast Routing Algorithms for LEO Satellite IP Networks," Chinese Journal of Aeronautics, vol. 20, no. 4, August, 2007, pp. 353-361.

[2] D. Yang, and W. Liao, "On multicast routing using rectilinear Steiner trees for LEO satellite networks," IEEE Transactions on Vehicular Technology, vol. 57, no. 4, July, 2008, pp.2560-2569.

[3] T. Mao, "A Multicast Routing Algorithm for LEO Satellite Networks," International Conference on Future Computer and Communication, June, 2009, pp.94-96.

[4] Z. Yin, L. Zhang, X. Zhou, P. Xu, and Y. Deng, "QoS-Guaranteed Secure Multicast Routing Protocol for Satellite IP Networks Using Hierarchical Architecture," International Journal of Communications Network and System Sciences, vol. 3, no. 4, 2010, pp. 355-363.

[5] S. Batti, M. Zghal, and N. Boudriga, "A New All-Optical Switching Node Including Virtual Memory and Synchronizer," JNW, vol. 5, no. 2, 2010, pp. 165-179.

[6] I. B. Djordjevic, and B. Vasic, "Novel combinatorial constructions of optical orthogonal codes for incoherent optical CDMA systems," Journal of Lightwave Technology, vol. 21, 2003, pp. 1869-1875.

[7] N. Boudriga, M. Sliti, and W. Abdallah, "Optical code-based filtering architecture for providing access control to all-optical networks (Invited)," International Conference on Transparent Optical Networks (ICTON), July, 2012, pp. 1-5.

 ISBN: 978-1-61208-317-9

DVB-S2 Extension : End-to-End Impact of Sharper Roll-Off Factor Over Satellite Link

Antoine Bonnaud and Eros Feltrin
Multimedia and Added Value Services Department
Eutelsat SA
Paris, France
e-mail: {abonnaud, efeltrin}@eutelsat.fr

Luca Barbiero
Department of Information Engineering
University of Padua
Italy
e-mail: barbier1@dei.unipd.it

***Abstract*—The world of the geo-synchronous satellite communications is witnessing, for several months now, a strong discussion, if not a dispute, between some systems manufacturers about the opportunity of increasing the efficiency of the coding schemes proposed by the second generation of the Digital Video Broadcasting standard by introducing some extensions. Even if the debate rises the interest of operators and manufacturers, its conclusions are far from being unanimously accepted. This paper concentrates mainly on the sharper low roll-off factor shaping filters proposed among the extensions. Their effects on the satellite channel are presented in a comparative analysis supported by measurements performed on commercial devices. The operational impacts are observed from the point of view of one of the main European satellite operators: Eutelsat.**

Keywords-DVB; satellite transponder; amplifier saturation

I. INTRODUCTION

In 2011, different manufacturers started to release new modulation and coding architectures for satellite communications, claiming to achieve significant performance improvements (higher than 15%) over their competitors and over the existing Second Generation of the Digital Video Broadcasting over Satellite (DVB-S2) standard [1]. The new features encompass, among others, sharp roll-off factors (ROF) for the shaping filters as well as proprietary pre-distortion and equalization techniques mitigating the non-linearity of the active components (mainly high power amplifiers).

These novelties arose the interest of the entire satellite communications world, including standardization committees, such as the current DVB-Technical Module, that are evaluating the insertion of these new features in the DVB-S2 standard [2].

Confronted with this innovation and the increasing requests for high efficiency in the capacity usage, Eutelsat, as one of the leading worldwide satellite operators, has to assume the role of customers guide in order to warrantee a high quality of service by using the very latest standard transmission techniques in the most correct way.

One of the most frequently asked questions concerns the real performance of ROF reduction over a satellite link in particular referring to the existing DVB-S2. The objective of the work presented in this paper is to assess by reliable and reproducible measurements in controlled laboratory environment the pros and cons of reducing the ROF in satellite transmission.

This study is based on measurements obtained with different professional devices and the close collaboration with key manufacturers in satellite communication industry. The devices under test are commercial products; the reported tests are performed over a reference DVB-S2 test bed [3].

The paper is organized as follows: Section II provides some backgrounds including the satellite channel description, the transmission techniques, their extensions and emission constraints. Section III is dedicated to a preliminary analysis of two operational scenarios. The validation equipment and methodology is described in Section IV, whereas the results are discussed in the Section V.

II. BACKGROUND

A. The Satellite Link

This paper considers operations over transparent geostationary satellites. To maximize their efficiency, the on-board amplifiers are operated as close as possible to the saturation point. This imposes a trade-off between power efficiency and signal distortion.

Further impairments onto the signal quality come from the Input Multiplexer (IMUX) before, and Output Multiplexer (OMUX) after the satellite Travelling Wave Tube Amplifier (TWTA) [4]. They aim at limiting interferences between adjacent transponders.

On the ground, the uplink earth station high power amplifier (HPA) output is not filtered. So proper back-off setting is of paramount importance to limit in-band and out-of-band distortion.

On the whole, satellite transmission is impacted by amplitude and phase distortions mainly due to amplifiers working in the saturation point, group delay, phase and frequency errors, Additive White Gaussian Noise (AWGN) and phase noise. This paper will consider the effect of amplifiers and AWGN only.

Studies on the non-linear degradation of end-to-end system performance exist in the literature and are here referred. These are simulations [5], extensive testing over emulator or satellite [6] and results obtained by the International Telecommunication Union Radio-communication Sector (ITU-R) [7].

 ISBN: 978-1-61208-317-9

B. Baseline Transmission Standard and Filtering

DVB-S2 [1] specifies the format of the digital signal to be transmitted over a satellite channel. It is used by the manufacturers as a reference document in the implementation of transmitters (or modulators) for broadcasting, interactive services, news gathering, and broadband applications that will be compatible with any receiver. The receiver design is not described in the DVB-S2 standard.

DVB-S2 introduces new modulation and coding (*modcod*) schemes with respect to the previous DVB-S. Three higher order modulation schemes, 8PSK, 16 and 32APSK. In particular, the last two suffer from non-linear distortions and their activation leads to a trade-off between power efficiency and signal degradation.

In DVB-S2, a conventional square-root raised cosine (RRC) shaping filter is defined at the transmitter. The 35% ROF is inherited from DVB-S whereas new values of 25% and 20% have been added. Today's industry manufacturers aim at extending the portfolio of their products by further deploying new *modcod*s and new ROF.

C. Extension Towards New Features and Sharper Filters

Sharpening the roll-off of the filters changes the characteristics of the modulated signal. The impact onto the entire transmission chain should be identified to derive the new operational parameters to be used, for ground and space segments, from consolidated experience.

The analysis and measurements performed by Eutelsat onto professional equipment is based on the following assumptions. The reference waveform is DVB-S2 with roll-off 35%. Competitive waveforms operate at sharper roll-off of 5% and 10%. The impact of compensation techniques (i.e., pre-distortion and equalization) is assessed but these additional techniques are considered as non-standard [1].

D. Eutelsat's Emission Constraints

Eutelsat aims at maintaining a high level of services and is mainly concerned about the amount of out-of-band emissions and the in-band distortion. That is why the European operator has published a standard, commonly known as EESS 502 [8], giving the basic requirements for the earth stations (E/S) and the spectrum utilization.

At the uplink E/S, the HPA sizing and linear operation has to be verified in order to reduce the risk of interference with other customers using the same Eutelsat satellite as well as with adjacent satellites.

At the receiver, the error performance over a non-linear channel should be known to ensure correct link closure, as well as the robustness against inter-channel interference (ICI). The two measurement points can be found at the HPA output and at the Low Noise Amplifier (LNA) input in Fig. 1.

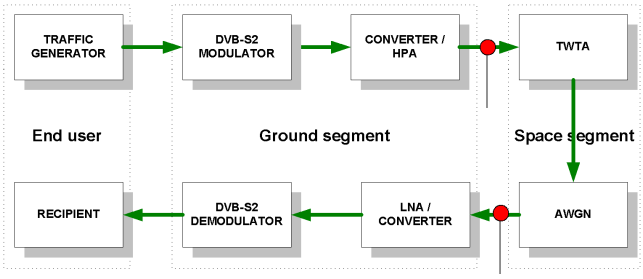


Figure 1. Emulated satellite chain and critical measurement: the IMUX and OMUX effects have not been investigated during the tests

III. PRELIMINARY ANALYSIS

A. Definitions

This section and the remaining part of the paper, considers the resources of satellite transponders, or capacity, in terms of bandwidth and power. As all natural resources, also the transponder bandwidth and power are limited: Eutelsat's satellites provide transponders with bandwidth of 33, 36, 72, 115 and 237.5 MHz whereas the power received on ground depends on the footprint extension.

In order to dissociate the power and the frequency, the power spectral density (PSD), or power per frequency unit (in dBW/Hz) is introduced. For linear modulation, the PSD is proportional, if not equivalent, to the symbol energy E_S which is representative of the quality of the received signal through the ration E_S/N_0 with N_0 the thermal noise energy.

From these definitions, it follows that a carrier modulated on a bandwidth B has a power C (in dBW) that is the mathematical integral of the signal PSD calculated over the bandwidth B at the receiver. Assuming an AWGN channel the noise power is given by $N = N_0 B$. The ratio C/N is also representative of the received signal quality and we can use the approximation

$$\frac{C}{N} = \frac{C}{N_0 \cdot B} \cong \frac{E_S}{N_0} \quad (1)$$

The expression (1) is already indicative of the fact that if we need to improve the performance of the received signal, the symbol energy must be increased. But, since the available power on board the satellite is limited to C, the only way is reducing the bandwidth B and increasing the PSD.

B. Two Reference Scenarios

The DVB-S2 (or related extensions) carriers can be used with two configurations: single and multiple carriers per transponder.

The first configuration is typical of consumer services like TV broadcasting and the forward channel in interactive networks (as it is described in the DVB-RCS standard [9]:

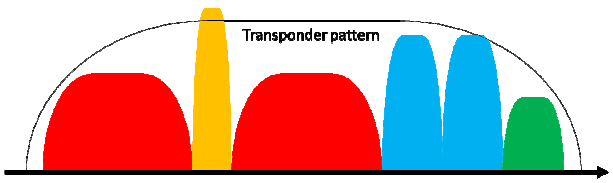


Figure 2. Multi-carriers per transponder configuration

 ISBN: 978-1-61208-317-9

one carrier exploits all the power available in a single transponder, also working close to the saturation point of the TWTA on board, and the signal coming from the satellite can be easily received also by small and cheap end user antennas.

With the same issue of exploiting the resources of a transponder, the configuration with only one carrier per transponder is used for the implementation of wideband data links, in which case the ground segment is represented by very large satellite dishes (typically over 3.2 m antennas). The entire transponder's power is required in order to reach the highest throughput by introducing high efficiency *modcod*s.

The second configuration, i.e., multiple carriers per transponder (see Fig. 2) is typical for the single channel per carrier (SCPC) adopted for relatively small data links based on Very Small Aperture Terminals (VSATs), video contributions and all inbound carriers of interactive systems.

The usage of the extended DVB-S2 features would bring important advantages to both these scenarios, but from the point of view of an operator there are also some counterparts to be taken into account.

C. Sharpening the Shaping Filters: Spectrum Analysis

First immediate advantage would come from reducing the ROF down to very low values. In a configuration with single carrier per transponder, this would permit to reduce the bandwidth occupied by the signal and increase the symbol rate and therefore the overall throughput.

For example, in a 36 MHz transponder, the maximum symbol rate reachable with a 20% ROF is $R_S = B/(1+ROF) = 30$ Msym/s, whereas with 5% ROF it would be 34.3 Msym/s. But also we have to maintain a constant value of the signal power because of the transponder limitation: decreasing the ROF means that the PSD must be reduced of the ratio $(1+20\%)/(1+5\%) = 1.14 = 0.6$ dB. The link budget would suffer from 0.6 dB reduction and, depending on the link margin assumed, this could impose to choose a less efficient *modcod* and the expected gain of 4.3 Msym/s would be hardly applicable to real cases. Certainly, this scenario would benefit from the introduction of new *modcod*s but only satellite links with large link margin would take a real advantage by the small ROF. In any case, this analysis does not take into account the distortion induced by the filters on both sides of the transponder: measurements are a helpful tool to study these effects, which could require a fine-tuning of the satellite link.

A similar analysis can be performed in the case of multi carriers per transponder where the impairment due to the ICI must be considered. As better shown in the following of this paper, the TWTA saturation generates a spectrum regrowth at the frequencies adjacent the signal band. Such a regrowth is more pronounced for small ROFs (see in particular Fig. 5) leading to increased ICI in a multi-carriers configuration. So, on one hand a small ROF would allow the reduction of the distance between the carriers, or carriers spacing, on the other hand this would also drive the transponders close to saturation, which would increase the interference.

Measurement results reported below in this paper will clarify this point and will permit to add further conclusions.

D. Low ROF Effects on the Constellations

One of the effects of sharpening the shaping filters appears in the plot of the transitions from one symbol to another within a constellation. These transitions generate a signal envelope variation over the maximum symbol energy which extent depends on the ROF: as smaller the ROF is as important is this envelope variation.

Fig. 3 shows the envelope variation at ROF 35% (black lines) and 5% (red) for two QPSK signals measured at the output of a modulator.

This behaviour has an immediate impact on the distortion and the Inter Symbol Interference (ISI) when the amplifiers are driven close to the saturation: as indicated in [5] pre-distortion techniques can be applied in order to mitigate this effect, even if they are less effective for high order constellations (APSK) than for QPSK and 8PSK. Small ROF would further reduce the pre-distortion performances. Also,

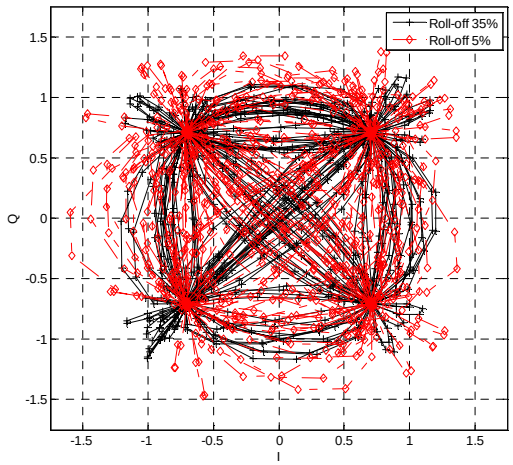


Figure 3. Plots of the QPSK constellation and transition between two symbols

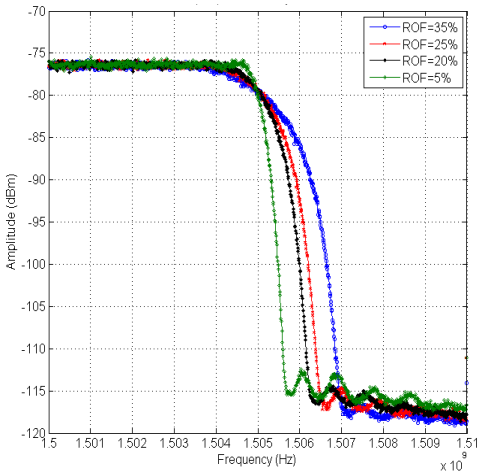


Figure 4. Sharper roll-off signal characterization: frequency response

 ISBN: 978-1-61208-317-9

this argumentation has been subject of some measurements described here below.

IV. THE MEASUREMENTS: THE TESTBED DESCRIPTION

A. The Channel Emulator

Eutelsat's validation equipment mainly consists of a reference satellite channel emulator [3] and state-of-the-art professional DVB-S2 transmitter and receiver. The entire DVB-S2 testbed also includes traffic generators and recipients, spectrum analyzers, a control unit, as well as additional DVB-S2 devices for a multi-carrier per transponder configuration.

The non-linear TWTA profile, defined in the DVB-S2 standard [1], Fig. H.12, is emulated in order to test the spectrum regrowth and the end-to-end performance over a non-linear channel in single carrier per transponder configuration. It is worth mentioning that the adopted profile is pessimistic compared with the modern TWTA characteristics. Anyway, the choice of this profile becomes mandatory in the case the studies done in the literature are taken as reference.

B. Test Conditions

A linear channel, i.e., a link with only the presence of AWGN, is often considered as a reference scenario to verify the performance of the channel decoder. Such a performance should be specified also in terms of difference between the minimum E_S/N_0 allowing a demodulator to lock a received signal (i.e., the lock threshold) and the E_S/N_0 at which the same device already locked to the signal passes to the unlock status (i.e., the unlock threshold).

For the emulation of a single carrier saturating a transponder, a memory-less non-linearity impairment on top of AWGN is added. A phase noise impairment is then introduced to represent the receive chain down-conversion characteristics (in the spacecraft, low-noise block down-converter, and tuner oscillators). At high R_S/B_T ratio (with R_S symbol rate, and B_T the satellite transponder bandwidth), a major impact is due to IMUX and OMUX filtering. In fact, since the signal works close to the transponder boundaries defined by these filters, a severe group delay affects the carrier: this aspect is not investigated hereafter.

Multiple carriers per transponder configuration consist in setting several carriers side by side, and insuring a limited adjacent channel interference of one to the other.

V. RESULTS

A. Transmitter End

Measurements at the modulator output and the HPA output have been performed in order to characterize the signal. At the modulator output the roll-off reduction causes sharper fall of the transmit filter frequency response (Fig. 4). The corresponding impulse response presents oscillations which amplitude decreases in time slowly than for high ROF filters. This behavior has its counterpart in the IQ signal diagram as already presented in Fig. 3.

At the output of the uplink high-power amplifier, the spectrum regrowth is given for Input Back-Off (IBO) values ranging from 0dB (saturation) to 18dB (linear region). Moving from 35% (Fig. 5-a) to 5% (Fig. 5–b) increases the out-of-band spectral regrowth by 1 to 2dB (depending on the IBO) as well as a change of the spectrum shape.

Therefore, whereas a sharper roll-off is activated with the aim of increasing the symbol rate (see the discussion above) a higher power is required to keep the E_S/N_0 ratio constant. Therefore, the uplink HPA could require a re-sizing in order to grant a level of unwanted out-of-band emissions compliant with the Eutelsat EESS 502 standard [8].

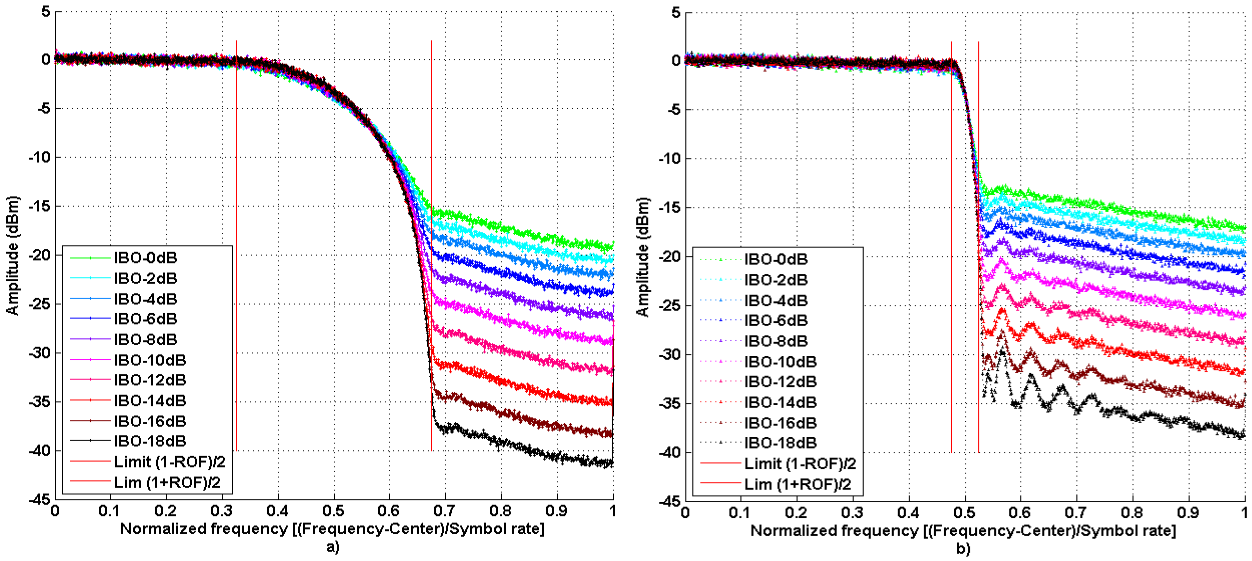


Figure 5. Spectrum regrowth versus IBO for roll-off 35% (a), and 5% (b)

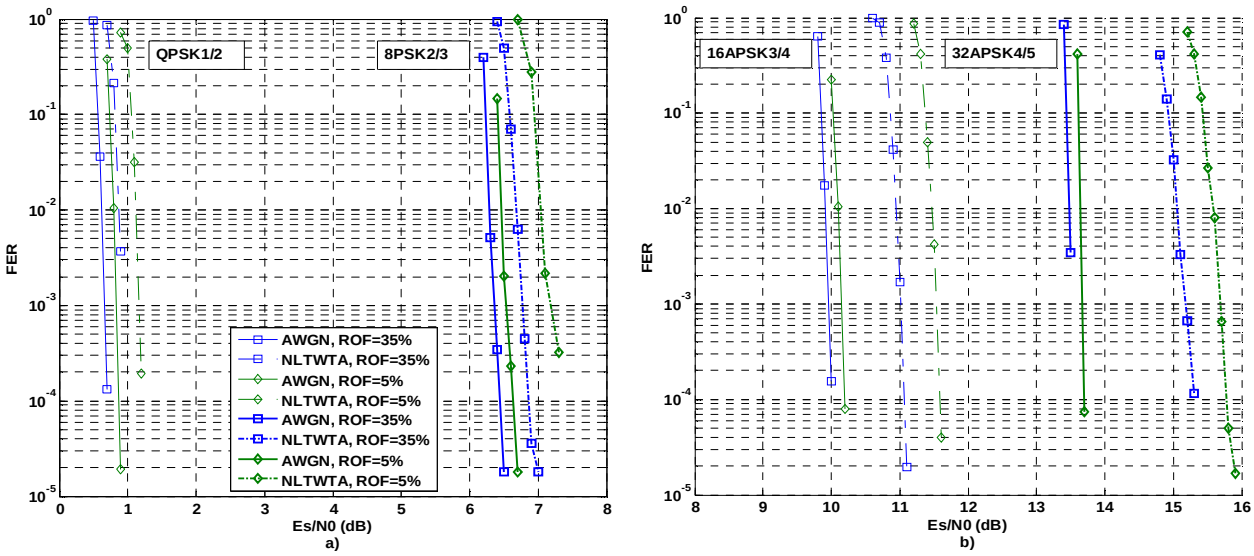


Figure 6. Impact of ROF reduction over linear and non-linear channel

In order to enable smaller ROF, E/S operators shall demonstrate that their out-of-band emissions are compliant with the Eutelsat's requirements. The same amount of spectrum regrowth is also present after on-board satellite TWTA, but will be partly rejected by the satellite OMUX filter.

B. *Receiver End: One Carrier per Transponder*

The error performance curves for *modcod*s QPSK 1/2 and 8PSK 2/3 are given in Fig. 6-a; those for 16APSK3/4 and 32APSK4/5 in Fig. 6-b. For every *modcod*, two channels are investigated (AWGN plain lines, and non-linearized TWTA dashed lines). For each channel the performance for ROF 35% (blue squares) and ROF 5% (green triangles) are presented.

In a linear channel, the most recent DVB-S2 demodulators overcome the first generation DVB-S2 chipsets by about 0.7dB to 1.7dB (from QPSK to 32APSK, see [7]). The use of ROF 5% always requires a E_S/N_0 ratio about 0.3dB higher to lock, but the steep waterfall behavior is unchanged. Over a non-linear channel at optimal IBO (given in [1], Table H.1) conventional ROF 35% suffers from low degradation compared to the reference figures. 5% ROF suffers from additional non-linear distortions of about 0.5dB.

For all *modcod*s, the expected frame error rate (*FER*) versus E_S/N_0 waterfall curves are shifted to a higher E_S/N_0 region when decreasing the ROF from 35% to 5% at the transmitter.

Reference simulations results presented in [5], indicate that the impact on the synchronization of the carrier of ROF 20%, 25% with reference to 35% is negligible. Our measurements onto professional equipment state that ROF 5% slightly degrades the error performance over linear and non-linear channels.

Another way to look at the non-linear channel is to search for the optimum back-off. The total degradation (D_{TOT}), as defined in [5] and reported in (2), is minimized when the sum of the non-linear distortions suffered by modulated signal and the power loss due to the operation of the amplifier at a given back-off is minimized.

$$D_{TOT} = \left[\frac{E_s}{N_0}\right]_{NL} - \left[\frac{E_s}{N_0}\right]_{AWGN} + OBO \qquad (2)$$

The optimum back-off values for ROF 35% and 5% are given in Fig. 7. ROF reduction to 5% does not change the output bock-off (OBO). The optimum IBO is larger, so is the total degradation.

So, when the ROF decreases, the loss over a nonlinear channel due to the distortion becomes stronger. To support the use of sharper roll-off, mitigation algorithms have been introduced. Some results of non-linear compensation techniques are presented in Fig. 7-b: even if the pre-distortion seems achieving better performance than equalization, it also suffers from several known limitations. First, TWTA characteristics must be known by the modulator in order to calculate the pre-distortion coefficients. Second, the pre-distortion algorithm minimizes the non-linear distortion at a given back-off. Consequently, at higher OBO the pre-compensation of the signal increases the degradation: in extreme conditions the usage of pre-distortion in linear region causes a huge penalty.

C. *Receiver End: Multiple Carriers per Transponder*

The measurements performed with ROF down to 5% highlight the risks for ICI due to the sharpening of the shaping filters. Fig. 8 is comparable to the Fig. 30 in [10] and shows that the degradation increases with the reduction of the carrier spacing (*CS)* much rapidly for very small ROF. The consequence is that, in particular for small carriers, the transmission is easily exposed to frequency errors that add heavy impairments due to ICI. In other words, very low tolerance is allowed to the central frequency for low ROF

 ISBN: 978-1-61208-317-9

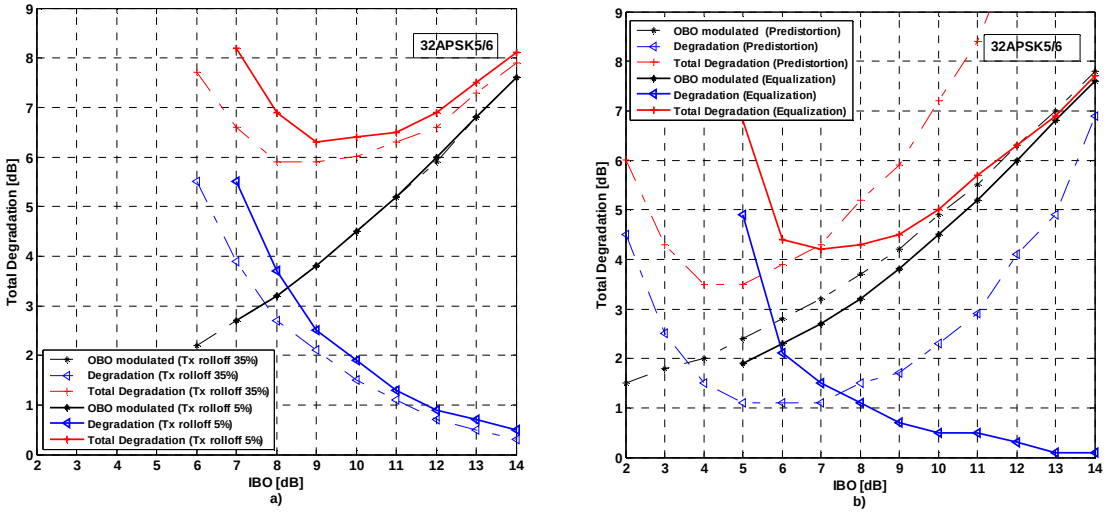


Figure 7. Signal degradation vs. IBO with ROF 35% and 5% a) and with pre-/post-compensation techniques b)

modulations. In the figure, the carrier spacing is normalized to the symbol rate R_S .

Another source of degradation of the end-to-end link appears when the receiver filter does not match the transmit filter. The dashed lines Fig. 8, labelled as "unknown" are an example of receivers using ROF higher - even if not explicitly indicated by the manufacturer - than the transmit filters. Obviously the ICI impact on the degradation is not negligible.

VI. CONCLUSIONS

This paper reports on the test results obtained by Eutelsat with professional DVB-S2 equipment featuring small ROF. Error performance in single and multiple carriers per transponder configurations have been assessed. The outcome unveils several issues as seen from a system operator perspective. At the receiver, the end-to-end link degradation is amplified. At the transmitter as well as on board satellite, non-linearity is increased. In multicarrier per transponder configuration, the old DVB-S2 only receivers suffer from additional losses.

In case of adoption of sharper roll-off factors, important changes must be planned at the transmitter and receiver. At the uplink station, more powerful high power amplifiers are required to limit out-of-band emissions.

At the modulator and demodulator ends, proprietary pre- and/or post-compensation techniques are available to mitigate the non-linear distortions. Their advantages and weaknesses have been highlighted, but since they are not part of a standard no commitment onto the interoperability and achievable performance among chipset manufacturers can be assured.

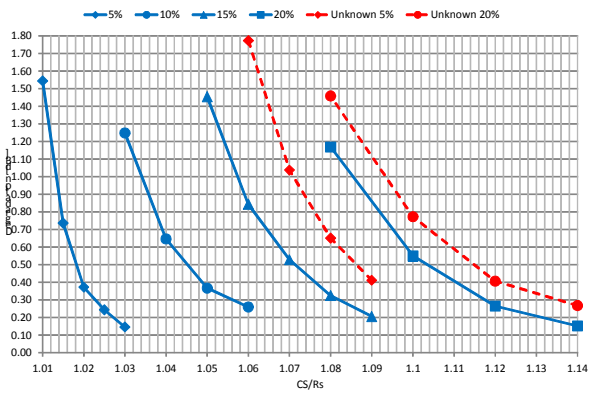


Figure 8. Degradation versus carrier spacing

REFERENCES

[1] ETSI EN 302 307 v1.2.1 (2009-08), "Digital Video Broadcasting (DVB); Second generation framing structure, channel coding and modulation systems for Broadcasting, Interactive Services, News Gathering and other broadband satellite applications (DVB-S2)".

[2] "Maintaining Excellence Enhancing a Great Standard", DVB Scene Issue 42, August 2013.

[3] H. Bischl et al. , "Adaptive coding and modulation for satellite broadband networks: From theory to practice" Int. J. Commun. Syst. Network, 2010, vol. 28, pp. 59–111.

[4] Teresa M. Braun, Satellite Communications Payload and Systems, 2012.

[5] E. Casini, R. De Gaudenzi, and A. Ginesi, "DVB-S2 modem algorithms design and performance over typical satellite channels"; Int. J. Satell. Commun. Networks , 2004; Vol. 22, pp.281-318.

[6] N. Girault, C. Moreau, and G. Dane, "DVB-S2 satellite experiment - Final Report", ESA project contract No. 19572/06/NL/JA.

[7] "Digital satellite broadcasting system (television, sound and data) with flexible configuration", Report ITU-R BO.2101, 2007.

[8] Eutelsat's Earth Station Minimum Technical and Operational Requirements, Standard M, EESS 502, Issue 14 Rev. 0, June 2011.

[9] ETSI TS 101 545-1 v1.1.1 (2012-05), "Digital Video Broadcasting (DVB); Second Generation DVB Interactive Satellite System (DVB-RCS2); Part 1: Overview and System Level specifications".

[10] ETSI TR 102 376 V1.1.1 (2005-02) (2012-05), "Digital Video Broadcasting (DVB), User guidelines for the second generation system for Broadcasting, Interactive Services, News Gathering and other broadband satellite applications", (DVB-S2)".

 ISBN: 978-1-61208-317-9

Application Layer Source-Channel Video Coding for Transmission with Smartphones over Satellite Channels

Igor Bisio, Aldo Grattarola,Fabio Lavagetto, Giulio Luzzati, Mario Marchese
DITEN - University of Genoa
Genoa, Italy
Email: igor.bisio, aldo.grattarola , fabio.lavagetto, giulio.luzzati, mario.marchese@unige.it

***Abstract*—Transmitting video with portable devices (such as smartphone), over heterogeneous networks with satellite portions, to a Remote Monitoring Host (RMH) represents a hot scientific and technical topic. Unfortunately, heterogeneity often implies impairments such as packet losses, due to errors and congestion, which negatively affect the video quality. To guarantee a satisfactory video fruition at the RMH, an application layer joint coding algorithm for video transmission is presented. It adaptively applies both video compression and encoding to protect the sent video stream, at the application layer, on the basis of the overall network condition estimated in terms of both maximum allowable throughput of the network and quality (packet cancellations or *lossiness*). A preliminarily performance study, carried out with real implementation of the algorithm, compares the joint coding against fixed schemes.**

***Keywords*—Application layer; joint source channel coding; disaster recovery.**

I. INTRODUCTION

The recent spreading of mobile computing platform such as those constituted by smartphones has provided grounds for ubiquitous connectivity. Much research has been done in the field, especially in the light of the recent advances in wireless communication [1], and [2]. The reference network considered in this paper is composed of heterogeneous portions, which include two main components: radio and satellite, which allows wide coverages. Offering real services with a specific guarantee of Quality of Service (QoS) over these hybrid systems is very challenging, as implies the solution of research and development issues due to the peculiarity of the application environment: efficient communication networks composed of mobile and fixed nodes operating with radio and satellite links. For this reason, the study and the tests in this paper is concentrated on the integrated wireless-satellite components of the network, where also much work has been done [3], [4]. These portions, unlike wired channels where information loss is mainly imputable to network or resource congestion, wireless media are characterized by low Signal-to-Noise Ratios (SNRs). This leads to Bit Error Rates (BERs), which amount may range from negligible to almost completely impairing. Moreover, when dealing with wireless channel the time invariance assumption can no longer hold: typical wireless channels may exhibit extremely quick dynamics, due to a number of factors, such as multipath fading, shadowing, radio interferences and weather conditions.

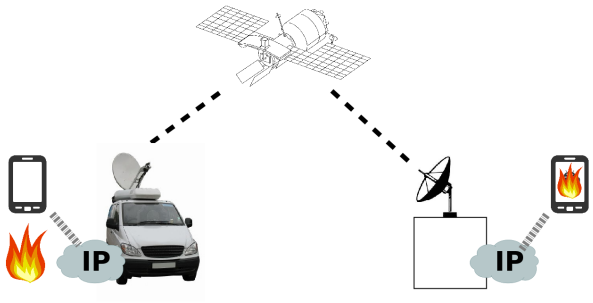


Fig. 1: The reference scenario

The specific scenario considered in this paper (reported in Fig.1) consider the case in which there is the need for transmitting a video stream over a satellite link. It often happens when an emergency link need to be established between an incident site and a RMH where a command center is operating. Unfortunately, heterogeneity often implies impairments such as packet losses, due to errors and congestion, which negatively affect the video quality. With this kind of scenario in mind, a static management of how the information is compressed and protected throughout its flow clearly constitutes a non optimal choice: one would want to be able to dynamically adapt the information flow by opportunely tuning the amount of information offered to the transmitting apparatus and how it is being protected, possibly jointly considering the impact each of these these tunings has on each other and on the whole system performance.

In more detail, to guarantee a satisfactory video fruition at the RMH, an application layer joint coding algorithm for video transmission is presented. It adaptively applies both video compression and encoding to protect the sent video stream, at the application layer, on the basis of the overall network condition estimated in terms of both maximum allowable throughput of the network and quality (packet cancellations or *lossiness*). By choosing to exploit the tools provided at the application layer, we develop an application layer joint coder completely oblivious of what happens at the lower network

 ISBN: 978-1-61208-317-9

layers, thus being able to be deployed on top of existing infrastructure without any need for rewiring or replacing hardware components. Application layer entities talk to each other via datagrams, so whenever one packet is sent, it is either delivered (possibly with some amount of delay) or lost as a whole. This can be modelled by means of a Packet Erasure Channel (PEC) model, and it is the reason we chose to protect information using a packet level channel code as detailed in the next Sections.

The remainder of the paper is organized as follows: Section II surveys the state of the art regarding the application layer coding and source-channel join coding problems; in Section III is described the proposed Application Layer Joint Coding (ALJC) approach. The simulative scenario developed to validate the proposed algorithm and the obtained numerical results are discussed in Section IV. The obtained results represent a initial study of the performance whose development has been proposed in [5]. Finally, the conclusions are drawn in Section V.

II. State of Art

In the literature, in particular in the field of satellite and space communications, it is argued that application-layer coding, obtained by applying redundancy at the application layer to protect the sent information, may be used to efficiently recover original data by guaranteeing flexibility and easy-configurability [6]. Nevertheless, the advantages of applying coding strategies at the application layer may improve the performance only in systems with low channel error rates because high error rates imply high levels of redundancy thus causing information losses due to congestion over a network. In more detail, starting from the Separation Theorem formalized by Shannon in [7] which is based on some assumptions such as infinite coding length, no constraints on complexity and on delay, in [8] the existence of two performance regions has been formally demonstrated. In a region, the employment of application layer coding is significantly advantageous while it is quite useless in the second. The first performance region is represented by systems that experience low level of intrinsic loss probability. On the contrary, applying such coding in high intrinsic loss probability systems is not advantageous because the high level of necessary redundancy causes a congestion growing. Differently, [9] proposed a slightly different point of view of the problem, whereas increasing protection does not directly result in increased offered load: in this sense, an end-to-end distortion minimization algorithm is devised. However, to the best of the authors' knowledge, literature lacks investigation upon actual implementation of a joint source-channel coding at the application layer, whereas proposed approaches take channel state knowledge for granted.

The aforementioned results inspired, as done similarly in [10], the rationale under this paper is that the boundary between the two regions can be moved so increasing the region size where the usage of the application layer coding is useful. In technical words, if the information sent from the sources is previously compressed the load offered to the network will be reduced. As a consequence, the redundancy employment does not cause a critical network load increasing.

For this reason, when dealing with practical radio networks it may be beneficial to handle compression and protection jointly, taking into account how one affects the other also. Moreover, differently from the aforementioned approach, our work introduces a method to estimate of the overall network condition estimated in terms of both maximum allowable throughput of the network and quality (packet cancellations or *lossiness*). It allows adapting, dynamically, compression and protection applied, in the case of this paper, to the transmitted video.

III. The Heuristic ALJC Approach

The technical reference of the proposed ALJC approach is reported in Fig.1. In the depicted scenario, video streams are transmitted, with portable devices (such as smartphone), over terrestrial/satellite networks, to a RMH. It may support emergency and rescue operations after crisis situations. For this reason, the practical implementation for the proposed ALJC has been developed on the *Android* operating system, by building two separate applications: a transmitter and a receiver. The chosen source encoding for the video frames is MJPEG due to its implementation and management easiness, as well as its intrinsic resiliency: it does not remove temporal correlation, so that losses and errors do not propagate in consequent frames, i.e. the stream does not lose synchronization even if the link temporarily drops. For what concerns the channel coding, LDPC [11] has been devised as a practical and computationally feasible solution. Finally, the stream of information is then handled to the UDP transport protocol.

Working on *Android* operating systems, we already have control of the source coder, by means of a JPEG encoding class which allows to specify a compression index q, in the $[0, 100]$ range, from worst to best quality, respectively. Considering the aforementioned optimization framework $s_{1k} = q$, $\forall k \in [1, K]$, $M = 1$. The chosen resolution for the video frames is the standardized resolution QCIF (176 x 144 pixels) .
The LDPC codec has been ported from an existing implementation [12]. The source code has been adapted and wrapped to be used as a native library within the *Android* Native Development Kit (NDK). It is a packet level LDPC code, meaning that it allows recovering of whole packets of data (i.e., a codeword is a set of packets). The encoder is used to dynamically set the employed code rate fec (i.e., the number of packets containing information over the overall number of packets composing a codeword). Considering again the aforementioned optimization framework $c_{1k} = fec$, $\forall k \in [1, K]$, $N = 1$.
The tests presented in this paper were carried out using packets 1024 bits long, which compose codewords 35 packets long. However, the LDPC mechanism needs some additional control information to be transmitted along with the source information flow, namely sequence numbers which identify codewords, packets within codewords and contents within the packets. This information is carried by a 24 bits additional header, leading what we call an *application layer packet* to be 1048 bits long.

 ISBN: 978-1-61208-317-9

The adaptiveness is made possible by means of a feedback (return) channel, that is, the receiver periodically sends information to the transmitter in the form of cumulative *ack* packets that carry information about how many packets have been lost, sent each time a new codeword is received, which allows the transmitter to make estimates about the overall network state. The return channel in our experimentation is error-free. In this sense, the key point for a system that aims at managing the source and channel coding jointly is to estimate how fast the hybrid terrestrial/satellite network is able to deliver the information and how vulnerable is the information in the process of traversing the entire network, in particular the radio/satellite portions. The estimation of the maximum allowable throughput of the network is made by using a particular packet buffer, which is filled by the LDPC encoder and it is emptied when an *ack* arrives (i.e., a packet is not removed after it has been sent, but it remains in the buffer until it is marked as *acked* or a *timeout* expires). The network service gross rate is given by the rate at which this buffer is being emptied.
On the other hand, since *ack* packets carry the lost packet count, provided by the LDPC decoder on the receiving side, we can use this amount as a measure of the quality (packet cancellations or *lossiness*).
The adaptive ALJC algorithm has the granularity of a codeword and, for each decision stage, it first assess the amount of protection needed to successfully traverse the entire hybrid network, then it tries to best exploit the remaining packets to try to fit in frames with the goal of maintaining a target frame rate expressed in Frame-Per-Second (FPS_{target}) and equal to, in this paper, 10 frames per second. To do this, we firstly tabulated the image size associated to each compression factor from 0 to 100 then inverted the function and computed a polynomial fitting. The mentioned actions have been performed in order to be able to obtain the compression quality needed as a function of the desired image (i.e., the single MJPEG video frame) file size.

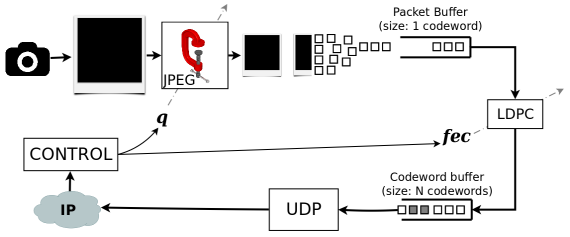


Fig. 2: The structure of the framework implementation

IV. Performance Investigation

A. Testbed

We realized a testbed to simulate the scenario depicted in the introduction. In this configuration, two separate *Android* devices communicate through a WiFi local network connected to a machine which emulates the effect of a satellite channel. On the receiving side, another WiFi network is used to interconnect the second device, which is supposed acting at the RMH.

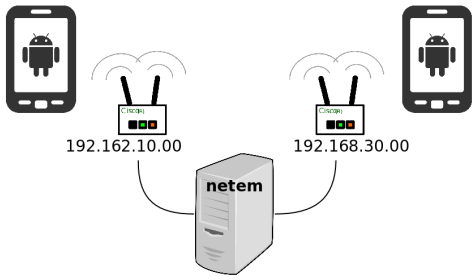


Fig. 3: The terrestrial/satellite network emulation apparatus used

This is reported in Fig.3, and it consists of a regular PC running with a Linux based operating system and equipped with two network interface cards, each connected to a WiFi access point. This way two separate sub-networks are created, and by using the *netem* tool it is possible to easily manage traffic going out of each interface, by tuning the bandwidth, the packet loss and bit error rate due to the emulated satellite channel, the delay and the packet buffer queuing policy.

B. Scenarios and Performance Metrics

TABLE I: The test scenarios

	Bandwidth	BER
A	400 Kbps	0 %
B	400 Kbps	10 %
C	180 Kbps	35 %
D	180 Kbps	0 %
E	180 Kbps	10 %
F	180 Kbps	35 %

TABLE II: The transmitter application source and channel coding settings

	FEC Rate	Compression Index
ALJC	*dynamic*	*dynamic*
Minimum Protection	30/35	60
Maximum Protection	4/35	20

In order to evaluate the behaviour of the ALJC system, we compared with two antipodal and static policies (i.e., minimal protection and maximum protection (see TABLE II). The compression index has been set to 20 in the case of maximum protection in order to provide a minimal intelligibility of the image (when equal to 0 the image information is almost completely lost), and to 60 because it conveys a frame image size that lies in the middle of the range. The test runs evaluate each coding choice behaviour during three minutes long sessions, exploiting the aforementioned network emulating machine that simulates different bandwidth and BER conditions experienced by the satellite link. A second run of tests deals with the system's adaptation capabilities in

 ISBN: 978-1-61208-317-9

time varying conditions.
We devised six plausible scenarios, listed and described accurately in Table I. In every scenario we had the network emulator introduce a 600 [ms] propagation delay to simulate a the presence of a geostationary satellite link in the network, as described in the introduction. In order to be able to evaluate and correctly understand how each coding policy actually behaves, we had to devise the key factors to measure in a video streaming framework. Intuitively, from a qualitative point of view it would be desirable for such a system to provide the highest possible image quality, as well as a fluent video reproduction, and naturally, a minimal information loss due to network impairments.

In order to measure the quality of individual frames of the MJPEG sequence, we utilize a well known metric, the Structural SIMilarity index. It was first introduced in [13], and it provides a quality measure of one of the images being compared, provided the other image is regarded as of perfect quality. The SSIM represents a good choice since it follows the MOS more closely than other indexes such as the Peak Signal to Noise Ratio (PSNR) or Mean Square Error (MSE). The SSIM index is computed over small patches of image, and the whole image index is obtained by averaging the individual patches' values. The individual SSIM value for a patch is given by

$$SSIM(x,y) = \frac{(2\mu_x\mu_y + C_1) + (2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \quad (1)$$

where x and y are the two small image blocks, μ_i is the i-th block pixel value average, σ_i is the i-th block pixel standard deviation, and

$$\sigma_{ij} = \frac{1}{M-1}\sum_{k=1}^{N}(i_k - \mu_i)(j_k - \mu_j) \quad (2)$$

where M is the number of pixel contained in the patch.

The SSIM index ranges from 0 (completely uncorrelated images) to 1 (identical images), and this is a useful feature since it can be considered a degradation factor.

In order to evaluate the performance of a given test run we devised a performance index with the following requirements

- it must reward high quality frames
- it has to reward a fluent video stream, i.e. a high frame throughput
- it must penalize corrupted or lost frames

We found that the following index S satisfies such requirements

$$S = \frac{\sum_{i=1}^{M} SSIM(f_i, \widehat{f}_i) \cdot f_{received}^{TOT}}{T_{sim}} \quad (3)$$

and can be interpreted as a *quality-weighted average framerate*.

For experimental purposes, in order to collect the information needed to compute such metrics, we arranged the implemented smartphone applications to write down each image frame: the transmitter writes a full quality version, while the receiver application records the source encoded version (i.e., including information loss due to compression). The frame written out by the smartphone applications are temporally referenced, so that it is possible to also infer information about the frame rate.

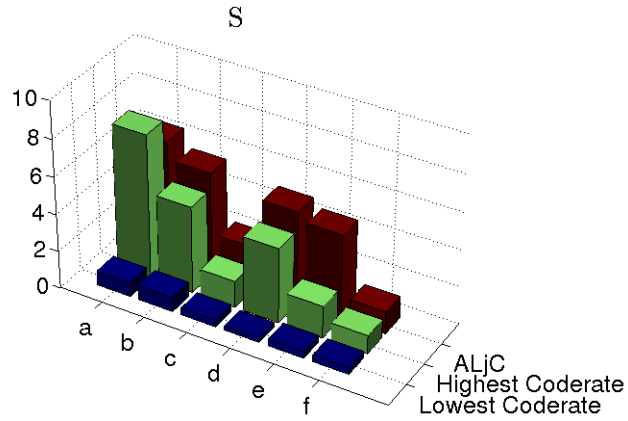


Fig. 4: Results of the simulation runs

A test run is structured in the following steps:

1) The smartphone starts streaming to the receiver, which is individuated through its IP address.
2) When the duration limit is reached, the transmitter stops broadcasting and sends to the server the reference file containing the original full quality frames.
3) As soon as it is done sending, the receiver sends its data, containing the received (and possibly corrupted) video frames.

C. Emulation Campaigns Results

In this section we show how our joint coding framework behaves in a stationary network condition (i.e., within channels whose characteristics do not vary over time), referring to Fig.4. As mentioned in Section IV-B, scenarios A, B and C, which set a bandwidth of 400 kbps and different BER conditions.

Scenario A simulates a wideband, error free channel: in this particular case the adaptive ALJC policy seems to perform slightly worse than a minimum protection one. This is reasonable, since the latter consistently employs high quality compression and waste a little quantity of bandwidth for protection, while ALJC more cautiously because the compression tuning "oscillates" due to a non-perfect network condition estimation in terms of both maximum allowable throughput of the network and quality. There is a minimal amount of lost packets to be imputable to kernel drops due to busy CPU. Naturally, the maximum protection approach here is the worst scoring, due to the unnecessarily poor quality of the frames and the low throughput due to high bandwidth waste. Scenario B is much more interesting: a significant BER (10%) causes the minimum protection policy to lose a substantial amount of packets, thus bringing down the overall sequence PSNR.
A very high BER (35%) represents a highly challenging

 ISBN: 978-1-61208-317-9

channel: here the performance index cannot be as high as in previous scenarios, because the ALJC policy needs to cut down on the offered load by using more aggressive compression in order to let the frames through the narrower window left by the channel encoder. At the same time, the minimal protection approach now loses nearly half the information fed into the network, while the conservative maximum protection policy is not able to transport a sufficient amount of information.

Scenarios D, E and F simulate narrowband channels instead. The behaviour is roughly the same as in the previous scenarios.

V. CONCLUSIONS

In this paper we introduced an application layer joint coding algorithm for video transmission. It allows the transmission of video streams on networks based on time varying and possibly lossy networks, such as those entailed by hybrid terrestial/satellite networks. It is shown that using only information available at the application layer, we can implement a system that outperforms oblivious static coding under nearly every network condition. From the practical viewpoint, it adaptively applies both video compression and encoding to protect the sent video stream, at the application layer, on the basis of the overall network condition estimated in terms of both maximum allowable throughput of the network and quality (packet cancellations or *lossiness*).

REFERENCES

[1] G. Araniti, M. Condoluci, A. Molinaro, A. Iera, and J. Cosmas, "Low complexity subgroup formation in lte systems," in *Broadband Multimedia Systems and Broadcasting (BMSB), 2013 IEEE International Symposium on*, 2013, pp. 1–6.

[2] L. Militano, M. Condoluci, G. Araniti, and A. Iera, "Multicast service delivery solutions in lte-advanced systems," in *Communications (ICC), 2013 IEEE International Conference on*, 2013, pp. 5954–5958.

[3] S. Mukherjee, M. De Sanctis, T. Rossi, E. Cianca, M. Ruggieri, and R. Prasad, "On the optimization of dvb-s2 links in ehf bands," in *Aerospace Conference, 2010 IEEE*, 2010, pp. 1–11.

[4] M. De Sanctis, E. Cianca, and M. Ruggieri, "Ip based routing algorithms for leo satellite networks in near-polar orbits," in *Aerospace Conference, 2003. Proceedings. 2003 IEEE*, vol. 3, 2003, pp. 3-1273–3-1280.

[5] I. Bisio, A. Grattarola, G. Luzzati, F. Lavagetto, and M. Marchese, "Performance evaluation of application layer joint coding for video transmission with smartphones over terrestrial/satellite emergency networks," in *International Communications Conference (ICC 2014), 2014 IEEE*, 2014, submitted.

[6] T. de Cola, H. Ernst, and M. Marchese, "Performance analysis of ccsds file delivery protocol and erasure coding techniques in deep space environments," *Comput. Netw.*, vol. 51, no. 14, pp. 4032–4049, Oct. 2007. [Online]. Available: http://dx.doi.org/10.1016/j.comnet.2007.04.015

[7] C. E. Shannon, "Coding theorems for a discrete source with a fidelity criterion," *Institute of Radio Engineers, International Convention Record*, vol. 7 (part 4), pp. 142–163, 1959.

[8] Y. Choi and P. Momcilovic, "On effectiveness of application-layer coding," *Information Theory, IEEE Transactions on*, vol. 57, no. 10, pp. 6673–6691, 2011.

[9] F. Zhai, Y. Eisenberg, T. Pappas, R. Berry, and A. Katsaggelos, "An integrated joint source-channel coding framework for video transmission over packet lossy networks," in *Image Processing, 2004. ICIP '04. 2004 International Conference on*, vol. 4, 2004, pp. 2531–2534 Vol. 4.

[10] I. Bisio, F. Lavagetto, and M. Marchese, "Application layer joint coding for image transmission over deep space channels," in *Global Telecommunications Conference (GLOBECOM 2011), 2011 IEEE*, 2011, pp. 1–6.

[11] R. Gallager, "Low-density parity-check codes," *Information Theory, IRE Transactions on*, vol. 8, no. 1, pp. 21–28, 1962.

[12] "Planete-bcast, inria, ldpc codes download page," http://planete-bcast.inrialpes.fr/article.php3?id_article=16.

[13] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *Image Processing, IEEE Transactions on*, vol. 13, no. 4, pp. 600–612, 2004.

 ISBN: 978-1-61208-317-9

Performance Evaluation of Network Selection Algorithms for Vertical Handover Procedures over Satellite/Terrestrial Mobile Networks

Igor Bisio, Carlo Braccini, Stefano Delucchi, Fabio Lavagetto, Mario Marchese
DITEN - University of Genoa
Genoa, Italy
Email: {igor.bisio, carlo.braccini, stefano.delucchi, fabio.lavagetto, mario.marchese}@unige.it

***Abstract*—The network selection is a decisional process aimed at determining the Radio Access Network (RAN) that a Mobile Node (MN) has to use and represents the core-function of the Vertical Handover procedure. In this paper, the authors propose a new algorithm for the network selection, called Dynamic-Technique for Order of Preference by Similarity to Ideal Solution (D-TOPSIS) that is a new formulation of the TOPSIS algorithm aimed at performing the same selection but requiring a minor number of operations and consequently reducing the time necessary to perform the selection. The main contribution of this work is the evaluation of the D-TOPSIS performance in a scenario where are available simultaneously a satellite network and Wi-Fi and WiMAX networks. The proposed approach is compared with other network selection methods found in the literature in two different cases: *i)* pedestrian, with a MN speed equal to 3 [m/s] and *ii)* vehicular, with a MN speed equal to 10 [m/s]. The numerical results show that D-TOPSIS assures a good performance as well as a limited execution time in both cases.**

Index Terms*—Network Selection, Vertical Handover, Multi Attribute Decision Making, Satellite Networks.

I. INTRODUCTION

Nowadays the multitude of heterogeneous Radio Access Networks (RANs) available can assure ubiquitous communications to Mobile Nodes (MNs) independently of their positions. Among them, the Satellite Networks are very important because they assure a vast coverage area without requiring any infrastructure or any Base Station or any Access Point, and a MN has only to be equipped by an opportune interface to use this technology. As a consequence they are able to support a large mobility even in environments where other technologies, such as Wi-Fi and Wi-Max, are not available. On the other hand the satellite network can assure worst performance with respect to the classical wireless networks in terms of a minor available capacity or an higher MN power consumption to maintain active the communication. So, it is important to define a mechanism for the cooperation of all the aforementioned networks that enables the MN to dynamically select to use the most appropriate one.

In this scenario, two concepts play a fundamental role: the vertical handover and the network selection. The first process is referred to the change of the RAN used by a MN. This change is vertical if the two networks, involved in the process, belong to two different technologies. The network selection is the decisional process that is in charge to select the RAN that a MN has to use. The two issues are tightly linked because if the selected RAN is not in use the MN has to perform an handover. As a consequence, an important requirement in order to not negatively impact the whole handover process, as reported in [1], is a stringent time constraint for the selection problem. For this reason the algorithm applied to solve the network selection problem has to limit its number of operations in order to reduce the time necessary to perform the selection.

Several works in the literature are focused on the network selection problem. A wild used algorithm is the so called **RSSI Based** (see [2] and [3] among the others), that belongs to the *mono-attribute* group. This algorithm is quite simple, characterized by a limited execution time but it assures poor results because of it considers only one metric during the selection process. In fact, this algorithm measures the Received Signal Strength Identifier (RSSI) from the Access Point (AP) of all the networks available and selects the one with the highest value. But it is true that there are also other metrics, such as available Capacity, Packet Loss Rate, Packet Delay, Energy (or Power) Consumption and Monetary Cost, that may be considered in the network selection problem. As a consequence better performance can be obtained by a different group of algorithms, called *multi-attribute*, because of they are able to consider several parameters (i.e., attributes) simultaneously, selecting the network that represents a compromise among the considered attributes. Four are the most diffused algorithms of this group: *i)* the **Simple Additive Weight** (SAW), defined in [4], [5] and [6] among the others, assigns a value, often called cost, to each *alternative* computed as the weighted sum of the normalized value of each considered *attribute*. *ii)* similarly the **Weighted Product Method** (WPM) calculates a cost for each network as the weighted product of the value of each considered attribute. *iii)* The **Fuzzy Logic Based** is derived from fuzzy set theory. Concerning the network selection many papers (such as [7], [8] and [9]) present a combination of the Fuzzy Logic theory with multi-attribute algorithms. *iv)* Finally the **Technique for Order of Preference by Similarity to Ideal Solution** (TOPSIS) algorithm selects the alternative (i.e., the RAN) that simultaneously minimizes and maximizes the distance respectively between the Positive Ideal Solution and the Negative Ideal Solution. This algorithm represents the

 ISBN: 978-1-61208-317-9

starting point and a standard version of the algorithm proposed by the authors.
Even if the mobile scenario is highly dynamic, some characteristics of the available RANs, evaluated during the network selection problem, remain constant until the MN is inside the related coverage areas, independently of the position of the MN. Such metrics are, for example, the monetary cost of the RAN, its level of security and the power consumption. In [10] the metrics (or parameters) are grouped in three categories: static, dynamic and semi-dynamic. A similar classification is applied in this paper and a new version of the Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS) algorithm, called Dynamic-TOPSIS (D-TOPSIS), aimed at obtaining the same selection results and at reducing the number of operations and consequently the execution time, is presented. The rest of the paper is organized as follows: in Section II are described the TOPSIS algorithm and the D-TOPSIS proposed by the authors. The simulative scenario developed is described in Section III, and the numerical results of the execution time and of the performance comparison are discussed in the same section. Finally, the conclusions are drawn in Section IV.

II. The proposed network selection algorithm

The proposed algorithm is a modification of the well known TOPSIS method that, tanks to its elasticity and versatility, can be applied not only to network selection as in [11], [12] and [13] but also in other telecommunication fields such as the sensor networks [14]. The new algorithm determines the same solution of the TOPSIS but assures a great computational complexity reduction. It is worth noticing that the definition of the D-TOPSIS is reported in [15]. Nevertheless, the algorithm formulation is here presented for the sake of completeness.

A. The TOPSIS algorithm

TOPSIS considers all the *alternatives* (i.e., all the available RANs) defined by the values assumed by the considered *attributes*. The *i-th alternative* is defined by the vector $A_i = (x_{i1}, \cdots, x_{ij} \cdots, x_{in})$ for $i \in [1, m]$ where n and m are respectively the number of *attributes* and the number of *alternatives*. As reported in [16], the TOPSIS approach can be geometrically modelled with m points in a n-dimensional space where it is possible to apply the Euclidean Norm to compute the distance between each alternative and a reference point.
This algorithm is composed of the following steps:

- Calculation of the weighted normalized attribute values (1):

$$v_{ij} = w_j \frac{x_{ij}}{\sqrt{\sum_{i=1}^{m} x_{ij}^2}} \tag{1}$$

  for each $i = 1, \cdots, m$ *alternative*, for each $j = 1, \cdots, n$ considered *attribute*. w_j is the weigh associated to the $j-th$ *attribute*, and the condition $\sum_{j=1}^{n} w_j = 1$ must be hold.
- Identification of the Positive (A^*) and the Negative (A^-) Ideal Solution, shortly PIS and NIS respectively, as reported in (2).

$$\begin{aligned} A^* &= (v_1^*, \cdots, v_j^* \cdots, v_n^*) = \\ &= \Big((\max_i v_{ij} | j \in J_1), (\min_i v_{ij} | j \in J_2) | i = 1, \cdots, m \Big) \\ A^- &= (v_1^-, \cdots, v_j^- \cdots, v_n^-) = \\ &\Big((\min_i v_{ij} | j \in J_1), (\max_i v_{ij} | j \in J_2) | i = 1, \cdots, m \Big) \end{aligned} \tag{2}$$

  where J_1 represents the set of positive attributes, that needs to be maximized, and J_2 represents the set of negative attributes that needs to be minimized.
- Calculation of the Separation Measures (SMs), to evaluate the distance between *alternatives* and ideal point (i.e., the Separation Measures) applying the Euclidean Norm, as reported in (3).

$$\begin{aligned} S_i^* &= \sqrt{\sum_{j=1}^{n} (v_{ij} - v_j^*)^2}; \quad for \quad i = 1, \cdots, m \\ S_i^- &= \sqrt{\sum_{j=1}^{n} (v_{ij} - v_j^-)^2}; \quad for \quad i = 1, \cdots, m \end{aligned} \tag{3}$$

- Calculation of the Similarity Index (SI) as $C_i^* = S_i^- / (S_i^- + S_i^*)$. The values are in the range $[0-1]$, where $C_i^* = 0$ if the *alternative* coincides with the NIS (i.e., $A_i = A^-$) and $C_i^* = 1$ if the *alternative* coincides with the PIS (i.e., $A_i = A^*$). It means that the best *alternative* is the one with the higher similarity index associated.

B. New Formulation of the TOPSIS Algorithm

The new formulation of the TOPSIS algorithm is called Dynamic-TOPSIS (D-TOPSIS) to highlight the fact that the decision at the generic step t, which coincides with a given instant, takes in consideration the decision performed in the previous step $t-1$. In more detail, the *attributes* at a generic step t used to evaluate the $i-th$ *alternative* are divided into two groups: the *static attributes* $\mathbf{s}_i(t)$, that maintain constant their values at each t step when the *alternative* is available, and the *dynamic attributes* $\mathbf{d}_i(t)$. So for the $i-th$ *alternative* it is true that $s_i(t) = s_i(t-1)$. Consequently, it is possible to modify the formulation of the TOPSIS algorithm as described in the following. Supposing the network selection performed, periodically, each T second (i.e., the selection at the step $t+1$ is taken T seconds after the selection at the step t) and defining the $i-th$ *alternative* as in (4):

$$\begin{aligned} &\mathbf{A}_i(t) = (\mathbf{s}_i(t-1), \mathbf{d}_i(t)) = (s_{i,1}(t-1), \cdots, s_{i,j}(t-1), \cdots, \\ &s_{i,n_s}(t-1), d_{i,j}(t), \cdots, d_{i,j}(t), \cdots, d_{i,n_d}(t)) \end{aligned} \tag{4}$$

 ISBN: 978-1-61208-317-9

where n_s and n_d represent respectively the number of static and dynamic *attributes*.

The basic idea of the D-TOPSIS is to apply the standard version of the TOPSIS algorithm each time the set of available RANs changes (i.e., the MN enters or leaves a new RAN). Then the results obtained are saved and reused each successive selection in which the D-TOPSIS is applied, until the set of available RANs does not change again (i.e., MN enters or leaves another RAN). In other words the following steps, necessary to perform the selection, are different depending on the number of available alternatives at the instant t ($m(t)$). In practice, if $m(t) = m(t-1)$ is employed the dynamic version of the algorithm, while, if this condition is not verified, the standard version of the algorithm is employed. To clarify the D-TOPSIS approach, the aforementioned distance is represented in Figure 1. It is worth noticing that in this case only two *attributes* are considered ($n = 2$) for the sake of simplicity but this choice is not a limitation. Both the *attributes* are positive (i.e., they must be maximized). One of them is static (s_1) and the other is dynamic (d_1). In Figure 1 is represented only the distance from the NIS, nevertheless similar conclusions can be drawn for the PIS. In Figure 1 the

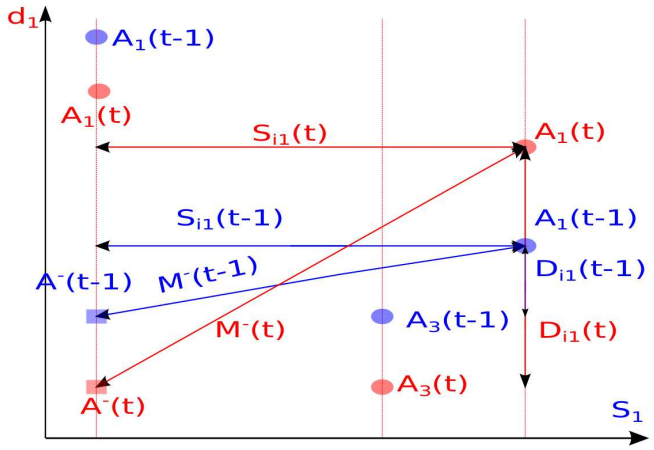


Fig. 1: Distance of the $i-th$ alternative from the Negative Ideal Solution (NIS)

NIS at the step t, $A^-(t)$ and $t-1$, $A^-(t-1)$, are identified by two squares, red and blue, respectively. The distance between the NIS and the $i-th$ *alternative* at the $t-th$ instant is determined by the two components $S_{i1}(t)$ and $D_{i1}(t)$, called in this work *Partial Distances*, that must be calculated for each *alternative* during each execution of the TOPSIS algorithm. It is possible to view in Figure 1 that $S_{i1}(t) = S_{i1}(t-1)$ and, consequently, it is not necessary to calculate the value of this parameter. It obviously reduces the number of the operation needed to carry out the RAN selection. Obviously this reduction becomes larger if the considered number of static parameters or the number of *alternatives* increase.

Starting from these considerations, and supposing that the number of available *alternatives* has not changed since the last network selection decision, the algorithm proposed is composed of the following phases:

- Calculation of the weighted normalized values of dynamic *attributes* only.
- Identification of the components of the Positive $A^*(t)$ and the Negative $A^-(t)$ Ideal Solution referred to the dynamic attributes. The other components of both vectors, referred to the static *attributes*, computed during the last standard TOPSIS execution, are constant, so it is not necessary to compute them again.
- Calculation of the *Partial Distances* for all the dynamic attributes between each *alternative* and the Ideal Solutions. This distance is defined as the square of the difference between the values assumed by the weighted normalized attribute and its ideal value. Also in this step the static attributes are not considered because their *Partial Distances* are equal to the distances computed during the last standard TOPSIS execution.
- Calculation of the Separation Measures (SMs) of each alternative, combining the *Partial Distances* of the dynamic attributes, calculated in the previous step and the *Partial Distances* of the static attributes, calculated in the last standard TOPSIS execution. The SM for a generic alternative is defined as the square root applied to the sum of the *Partial Distances* of all its attributes.
- Calculation of the Similarity Index as defined in the standard TOPSIS formulation. Clearly, also for the D-TOPSIS the highest Similarity Index identifies the best alternative.

Starting from the described phases, the reduction in the number of necessary operations is obvious. In practice, applying the D-TOPSIS for each *alternative*, it is not necessary to calculate both values of the *Partial Distances* for the static *attributes*. These values are stored during the last execution of the standard TOPSIS and simply loaded during the execution of the D-TOPSIS. Only the *Partial Distances* referred to the dynamic attributes must be calculated in order to compute the Separation Measures and the Similarity Index.

III. Performance Investigation

A. The Reference Scenario

The scenario taken as a reference in this paper is composed by a remote host that communicates with a MN. Two different MN speeds are considered: 3 [m/s] (pedestrian mobility case) and 10 [m/s] (vehicular mobility case). The communication is constituted by a UDP traffic flow that is transmitted by a remote host to a MN. Three different types of RANs are considered: Satellite Network, with a single footprint that covers the whole area, Wi-MAX, with again a single coverage area and eight Wi-Fi areas. The network selection is performed periodically each $T = 5$ [s]. Each simulation duration is set equal to 500 [s].

The metrics considered during the network selection are four: the Received Signal Strength Indicator, (*RSSI*), measured by the MN, the Capacity that the network reserves to the MN (*C*), the Monetary Cost (*MC*) paid by the user to use the network

 ISBN: 978-1-61208-317-9

and the Power Consumption (*P*) of the MN. Obviously the first two *attributes* are positive while the others are negative. It is worth noticing that, according with [10], *C*, *MC* and *P* are static *attributes* while *RSSI* is a dynamic *attribute*, computed as a function of the distance between the MN and the AP of the RAN in use for all the RAN except for the satellite network in which it is kept constant in all the cell. So $n_s = 3$ and $n_d = 1$.

Nine network selection algorithms are considered: five multi-attribute approaches, among them TOPSIS and D-TOPSIS are included. The other four are representative of each group presented in Section I. The other four are single-attribute, each of them is focused only on optimizing one of the considered attributes: *i)* Received Signal Strength Indicator based ($RSSI_b$), *ii)* Capacity based (C_b), *iii)* Monetary Cost based (MC_b) and *iv)* Power Consumption based (P_b).

The simulated scenario adopted has been developed through the Network Simulator 2 (*ns-2*). The dimension and the position of the coverage areas of each access network are randomly set in each simulation. Likewise the values of the considered attributes vary in each executed simulation, over the range reported in Table I. The Monetary Cost is modelled

TABLE I: RANGE VALUE OF THE ATTRIBUTE CONSIDERED

Parameters	Range Value
Power Consumption	[0,16 - 0,22] w
Monetary Cost	[1-10]
Wi-Fi Capacity	[1 - 20] Mbps
Wi-Max Capacity	2 Mbps
Satellite Network Capacity	0.384 Mbps

as an indicative number that ranks the network from the cheapest (*MC* $= 1$), to the most expensive (*MC* $= 10$) while the satellite capacity is obtained considering a bandwidth equal to $200[KHz]$, $C/N = 5.7[dB]$.

The numerical results obtained through a simulative campaign are grouped into two sets. The first one is aimed at highlighting the improvement assured by the D-TOPSIS, with respect to the standard TOPSIS and also with respect to the other algorithms, in terms of execution time. The second set reports a comparison between the performance obtained by the two TOPSIS versions that converge on the same solution and by the aforementioned algorithms.

B. Execution Time Comparison

Figure 2 shows the execution time for all the considered network selection algorithms for a different number of RANs, from two, the minimum number of *alternatives* to perform the network selection, to ten, the total number of RANs in the simulated scenario. D-TOPSIS is the second fastest algorithm among the considered ones; only the single-attribute method is faster. On the other hand these algorithms give poor results and a sub-optimal selection as highlighted in the next subsection. The execution time of the standard TOPSIS has intermediate performance while the fuzzy algorithm is the slowest algorithm because it implements several operation during the so called *fuzzyfication* and the *defuzzyfication* phases. So it is clear the advantage assured by the D-TOPSIS in terms of execution time, not only with respect to the standard TOPSIS but also with respect to all the other multi-attribute network selection algorithms.

Finally, it is possible to note that the difference between the execution times increases if the number of *alternatives* increases.

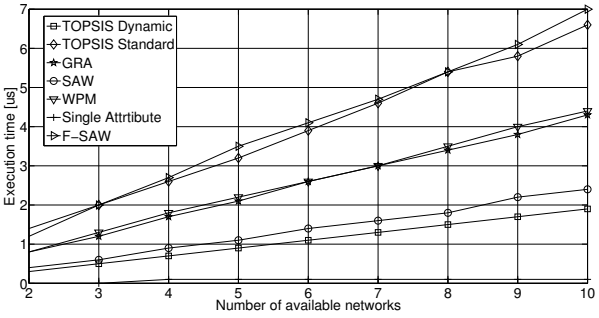


Fig. 2: Distance of the $i-th$ *alternative* from the Negative Ideal Solution

C. Performance Comparison

In this subsection, the performance obtained with the aforementioned algorithms are compared. As it is possible to see in Figure 3, the value of the metric H is much higher when the MN moves at a speed of 10 m/s with respect to when it moves at 3 m/s. This occurs because the MN is inside of each RAN for less time if its speed increases, and the scenario changes its characteristics more quickly.

Observing the power received by the MN, plotted in Figure 4, expressed by the metrics RSSI, it is possible to observe that also in this case the best results are obtained by pedestrian mobility. In fact, the permanence time of the node inside the Wi-Fi or WiMAX cells is higher for the pedestrian pace with respect to the case of vehicular mobility. If the node speed increases, it is necessary to use satellite networks more frequently which provide worse performance but at the same time more ample coverage.

Similar considerations can be made for the values assumed for some other metrics, especially the allocated capacity C (Figure 5) and the power consumption of the MN P (Figure 6). As a matter of fact also in this case a wider use of satellite networks corresponds to a worsening in the values assumed by the two metrics we have considered in this paper.

Different behaviour, characterized the delay of the traffic flows transmitted in downlink (D) and the monetary cost (MC) paid by the mobile node to use the network, plotted respectively in Figure 7 and in Figure 8. It is possible to observe similar performance for the two mobility types considered. In fact, the use of satellite networks guarantees minor allocated capacity at the terminal, but this reduction does not particularly affect the perceived delay of traffic flow, because the allocated capacity is sufficient to guarantee satisfying average performance. As far as monetary cost is concerned, we have decided to impose

 ISBN: 978-1-61208-317-9

the cost in the same range for both the satellite network and the other networks we have considered in this paper.
It is worth noticing that the comparison between the two mobility types is not fair because the MN in the vehicular mobility follows a longer path with respect to the pedestrian mobility. As a consequence when the speed is 10 m/s the MN stays outside the Wi-Fi and WiMAX coverage area for a longer time, and, consequently, it has to use the satellite network. Nevertheless, the performance comparison among the network selection algorithm is fear in both the considered mobility scenarios.
Considering the numerical results obtained through the conducted simulation campaign, one can observe that with a MN speed equal to 3 m/s the multi-attribute algorithms assure best performance with respect to the single-attribute ones even if these algorithms determine a minor number of handover executions (H) and a minor execution time. Moreover, it is possible to view that among the multi-attribute algorithms the best performance are obtained with both the TOPSIS algorithms that perform the same selection, choosing the RAN that assure the best compromise between the adopted attributes.
Similar results can be obtained considering the MN speed equal to 10 ms. A slight difference is that in this case also the SAW algorithm, and not only the TOPSIS algorithm, assure the best compromise among the considered performance metrics.

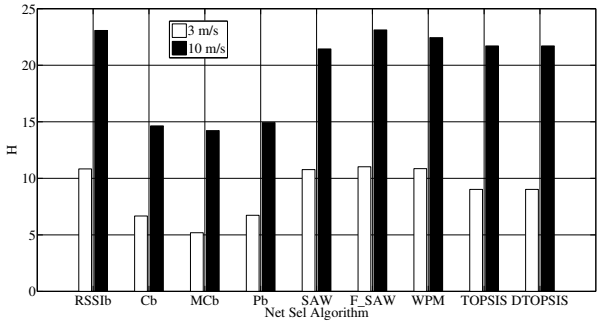


Fig. 3: Values of the H metric for different network selection algorithms for the two MN speed considered.

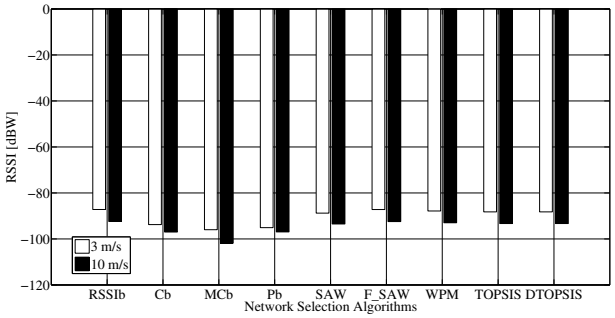


Fig. 4: Values of the $RSSI$ metric for different network selection algorithms for the two MN speed considered.

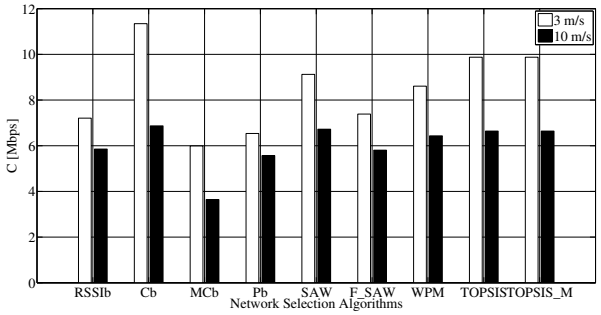


Fig. 5: Values of the C metric for different network selection algorithms for the two MN speed considered.

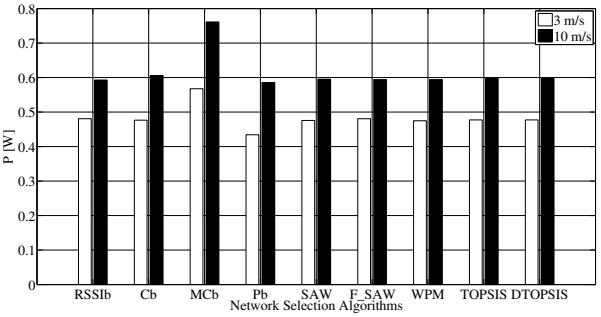


Fig. 6: Values of the P metric for different network selection algorithms for the two MN speed considered.

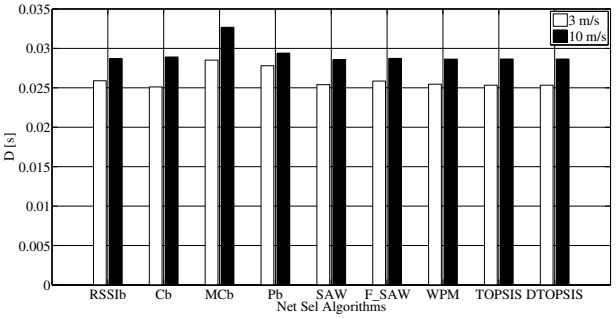


Fig. 7: Values of the D metric for different network selection algorithms for the two MN speed considered.

IV. Conclusions

A stringent time limit requirement in the network selection process is necessary to optimize the handover execution. Consequently, the network selection algorithms should be designed by limiting the number of operations needed to perform the selection. Starting from this necessity, a new version of the TOPSIS algorithm applied to the network selection problem, called D-TOPSIS, is proposed in this paper. It is aimed at determining the same selections but with a large reduction of the number of required operations.
The performance obtained by using D-TOPSIS are compared with the performance obtained by using the most important network selection policies found in the literature. Two mobility

 ISBN: 978-1-61208-317-9

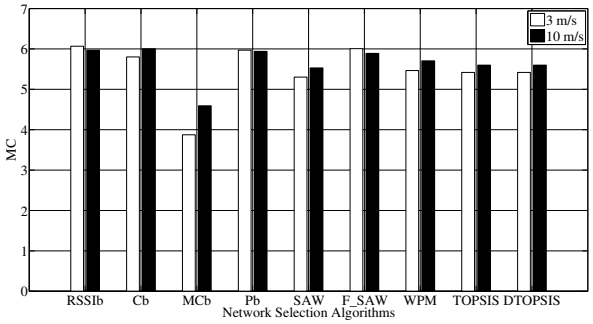


Fig. 8: Values of the MC metric for different network selection algorithms for the two MN speed considered.

scenarios have been considered: the pedestrian and the vehicular case. The numerical results demonstrate that the TOPSIS based approaches improve the performance of the network selection process and, in particular, D-TOPSIS reduces the execution time needed to carry out the network selection.

REFERENCES

[1] I. Chantaksinopas, P. Oothongsap, and A. Prayote, "Network selection delay comparison of network selection techniques for safety applications on VANET", Network Operations and Management Symposium (APNOMS), 2011 13th Asia-Pacific, 2011, pp. 1–7.

[2] S. Mohanty and I. Akyildiz, "A cross-layer (layer 2 + 3) handoff management protocol for next-generation wireless systems", Mobile Computing, IEEE Transactions on, vol. 5, no. 10, 2006, pp. 1347–1360.

[3] G. Chen, M. Song, Y. Zhang, X. Wang, and B. Sun, "End-to-end QoS guaranteed seamless handover scheme for cognitive heterogeneous networks", in Network Infrastructure and Digital Content, 2010 2nd IEEE International Conference on, 2010, pp. 440–444.

[4] W. Shen and Q.-A. Zeng, "Cost-function-based network selection strategy in integrated wireless and mobile networks", Vehicular Technology, IEEE Transactions on, vol. 57, no. 6, Nov. 2008, pp. 3778–3788.

[5] A. Hasswa, N. Nasser, and H. Hassanein, "Tramcar: A context-aware cross-layer architecture for next generation heterogeneous wireless networks", in Communications, 2006. ICC 06. IEEE International Conference on, vol. 1, 2006, pp. 240–245.

[6] R. Tawil, G. Pujolle, and O. Salazar, "A vertical handoff decision scheme in heterogeneous wireless systems", in Vehicular Technology Conference, 2008. VTC Spring 2008. IEEE, 2008, pp. 2626–2630.

[7] Q. He, G. Chen, and L. Zhang, "A vertical handoff decision algorithm based on fuzzy control in WiMAX and TD-SCDMA heterogeneous wireless networks", in Wireless Communications Networking and Mobile Computing (WiCOM), 2010 6th International Conference on, sept. 2010, pp. 1–4.

[8] J. Hou and D. OBrien, "Vertical handover-decision-making algorithm using fuzzy logic for the integrated radio-and-ow system", Wireless Communications, IEEE Transactions on, vol. 5, no. 1, 2006, pp. 176–185.

[9] Q. He, "A fuzzy logic based vertical handoff decision algorithm between WWAN and WLAN", in Networking and Digital Society (ICNDS), 2010 2nd International Conference on, vol. 2, May 2010, pp. 561–564.

[10] L. Wang and G.-S. Kuo, "Mathematical modeling for network selection in heterogeneous wireless networks: a tutorial," Communications Surveys Tutorials, IEEE, vol. 15, no. 1, 2013, pp. 271–292.

[11] I. Bisio, S. Delucchi, F. Lavagetto, and M. Marchese, "Simulative Performance Comparison of Network Selection Algorithms in the Framework of the 802.21 Standard," in Performance Evaluation of Computer and Telecommunication Systems (SPECTS), 2013 International Symposium on. IEEE, Jul. 2013, pp. 92–99.

[12] M. Lahby, L. Cherkaoui, and A. Adib, "Network selection algorithm based on DIFF-AHP and TOPSIS in heterogeneous wireless networks," in Multimedia Computing and Systems (ICMCS), 2012 International Conference on, 2012, pp. 485–490.

[13] A. Sgora, C. A. Gizelis, and D. D. Vergados, "Network selection in a WiMAXWiFi environment," Pervasive and Mobile Computing, vol. 7, no. 5, 2011, pp. 584–594.

[14] I. Bisio and M. Marchese, "Attributes definitions and measurement methods for madm based sink selection controls in satellite sensor networks," in Global Telecommunications Conference, 2008. IEEE GLOBECOM 2008. IEEE, 2008, pp. 1–5.

[15] I. Bisio, C. Braccini, S. Delucchi, F. Lavagetto, and M. Marchese, "Dynamic multi-attribute network selection algorithm for vertical handover procedures over mobile ad hoc networks," in Communications (ICC), 2014 IEEE International Conference on, 2014, submitted.

[16] P. K. Yoon, C.-L. Hwang, and K. Yoon, "Multiple Attribute Decision Making: An Introduction (Quantitative Applications in the Social Sciences)". Sage Pubn Inc, Mar. 1995.

 ISBN: 978-1-61208-317-9

Network Coded TCP (CTCP) Performance over Satellite Networks

Jason Cloud and Muriel Médard
Research Laboratory of Electronics
Massachusetts Institute of Technology
Cambridge, MA USA
email: {jcloud,medard}@mit.edu

Douglas Leith
Hamilton Institute
National University of Ireland Maynooth
Co. Kildare, Ireland
email: doug.leith@nuim.ie

***Abstract*—We show preliminary results for the performance of Network Coded TCP (CTCP) over large latency networks. While CTCP performs very well in networks with relatively short RTT, the slow-start mechanism currently employed does not adequately fill the available bandwidth when the RTT is large. Regardless, we show that CTCP still outperforms current TCP variants (i.e., Cubic TCP and Hybla TCP) for high packet loss rates (e.g., $> 2.5\%$). We then explore the possibility of a modified congestion control mechanism based off of H-TCP that opens the congestion window quickly to overcome the challenges of large latency networks. Preliminary results are provided that show the combination of network coding with an appropriate congestion control algorithm can provide gains on the order of 20 times that of existing TCP variants. Finally, we provide a discussion of the future work needed to increase CTCP's performance in these networks.**

***Keywords*—Network Coding; TCP; High Delay.**

I. Introduction

It is widely known that TCP performs poorly over satellite networks [1], [2]. The combination of long round-trip times (RTT) and high packet loss rates (PER) over these networks create an environment that seriously degrades the performance of TCP. To overcome the challenges presented by satellite communication, a large variety of solutions have been proposed over the years. These range from modifications to TCP's congestion control algorithm to implementing performance enhancing proxies (PEPs). Each, of which, usually have their own drawbacks. In the case of modified TCP protocols, adoption is prevented due to the specialized nature of the protocol and issues related to fairness with other TCP variants. In the case of PEPs, increased hardware costs and issues regarding end-to-end semantics is an issue. In this paper, we suggest the use of Coded TCP (CTCP) proposed by Kim *et. al.*, [3], to overcome a large number of these issues.

Providing reliable data transport for satellite environments has been a topic of study since the late 1990's [1], [2]. End-to-end solutions typically involve tuning TCP so that the long RTTs representative of satellite links do not negatively impact performance. Two versions that perform well over satellite networks are TCP Cubic [4] and TCP Hybla [5]. Cubic, designed for high speed networks, and Hybla, designed for heterogeneous networks, use a congestion window algorithm that increases the congestion window size ($cwnd$) independently from the RTT. This makes either version useful in environments with high delay. Unlike TCP Cubic, Hybla was developed to also reduce the impact of multiple losses, inappropriate timeouts, and burstiness making it a more logical option for use over satellite links. When compared with each other, studies have shown that Hybla performs better than Cubic under high PERs while the reverse is true under low PER's [6]. Regardless, both experience performance degradation under high losses. A more recent protocol, Loss-Tolerant TCP (LT-TCP) [7], [8], combines Reed-Solomon (RS) coding with TCP to overcome this issue, but it requires the use of explicit congestion control (ECN) and the RS code can result in performance loss due to decoding errors. CTCP circumvents these issues by using a congestion control algorithm that does not rely on feedback from lower layers and network coding eliminates the possibility of decoding errors while helping to overcome packet losses.

In lieu of changes to TCP, PEPs are another common method used to increase performance over satellite links. A TCP flow is generally terminated at the gateway to the satellite link, a protocol specifically designed for the satellite system (usually one that is proprietary) is used to transmit data over the satellite network, and a new TCP session is setup on the other side of the satellite link to complete the connection. This implementation poses two issues. First, the cost of implementing a PEP at the satellite network gateway may be high. Second, the termination of TCP sessions at the PEP violates end-to-end semantics such as IPSEC [2]. Again, CTCP has the potential to eliminate the need for PEPs while providing the same level of service over satellite links.

In this paper, we explore CTCP's performance in high RTT environments to determine if we can provide resilience in the presence of high packet loss rates, in addition to achieving the performance of TCP Cubic or TCP Hybla under no packet losses. We first provide a description in Section II of the existing, well tested version of CTCP that uses a TCP Reno style slow-start mechanism. This version is designed to provide robustness to packet losses through the use of network coding, but the congestion window management is ill-suited to large bandwidth-delay products (BDP). In Section III we measure the performance of a modified version of CTCP that opens $cwnd$ in a manner similar to H-TCP [9] to show that it is indeed possible to achieve high performance with large PER's and RTT's. Finally, we conclude in Section IV by proposing areas of possible future work.

II. CTCP Overview

The development of CTCP [3] has shown how the integration of network coding with TCP can provide significant

 ISBN: 978-1-61208-317-9

benefits over existing TCP variants, especially in high packet loss environments. These gains are a direct result of the combination of both network coding and CTCP's congestion window management. The remainder of this section will provide a brief introduction into both of these mechanisms. For a more detailed explanation of CTCP's implementation and performance, the reader should refer to [3].

A. Network Coding

One of the key features of CTCP is the use of network coding to aid in recovery from packet losses and the capability to decrease overhead by limiting the number of required retransmissions. The gains provided by network coding are twofold: network coded packets can be used to provide forward error correction in the case of lost packets, and also simplifies feedback and retransmissions (should they be needed). In its current implementation, CTCP uses a systematic random linear code [10]. As an example, consider the transfer of packets $p_1 \dots p_k$ between a server and client. Each packet $p_i, i \in \{1, \dots, k\}$ is first sent uncoded followed by a number of network coded packets where every coded packet c_i is a random linear combination of the packets $p_1, \dots, p_k$, i.e.,

$$c_i = \sum_{j=1}^{k} \alpha_j p_j, \tag{1}$$

each $\alpha_j \in \mathbb{F}_{2^q}$ is randomly chosen, and q is large enough to ensure linear independence among all network coded packets with high probability (the current implementation draws α_j from $\mathbb{F}_{256}$). Should retransmissions be needed, additional network coded packets are generated and sent to the client. The number of coded packets sent along with the uncoded packets is dynamically determined based on an estimate of the path's packet loss probability, while the number of packets sent as a result of feedback is determined by both the number degrees of freedom ($dofs$) required by the client to decode and the estimated packet loss probability.

B. Congestion Control

The second feature of CTCP that is a major contributor to the observed gains is the congestion window management. CTCP uses a modified version of TCP's Additive Increase, Multiplicative Decrease (AIMD) algorithm that was designed to be compatible with network coding. Specifically, the current implementation modifies the multiplicative back-off factor, β, to be

$$\beta = \frac{RTT_{\min}}{RTT}, \tag{2}$$

where $RTT_{\min}$ is the path's estimated true round-trip propagation delay (which is assumed to be the lowest per-packet RTT observed during the lifetime of a connection) and RTT is the last measured round-trip time. The congestion window is increased using TCP Reno's increase mechanism (i.e., the slow-start mode increases $cwnd$ by 1 for every received acknowledgement, otherwise $cwnd$ is increased by $1/cwnd$).

This approach, in effect, assumes that the increase of a packet's RTT is solely due to the queuing of packets along the path, which is an indication of congestion. If a packet is lost at random and $RTT = RTT_{\min}$ (i.e., it is not lost due to congestion), then $cwnd$ is not reduced. On the other hand if $RTT > RTT_{\min}$, a packet loss is interpreted as congestion and $cwnd$ is reduced by a factor of β. While this approach does a fairly good job at distinguishing between packet losses due to congestion and poor channels, it has a few characteristics that may not work well for satellite networks. A more detailed discussion is provided in Section IV.

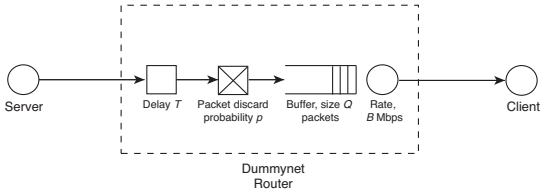


Figure 1: Schematic of experimental testbed.

C. Performance over Short RTT Networks

Using the mechanisms summarized above, [3] implemented CTCP as a SOCKSv5 proxy in user space and measured its performance over a wide range of conditions (although all measurements were made using round-trip times representative of terrestrial networks). Kim *et. al.* showed that CTCP can achieve goodput efficiencies greater than 90% for packet loss rates less than 0.2 while the performance of standard TCP variants is severely impacted. Another important aspect of CTCP is that it is friendly/fair with standard TCP, unlike some TCP variants that work well over satellite networks but are unfriendly to other TCP variants (e.g., TCP Hybla and Cubic TCP [11]). This is important since we are interested in providing an end-to-end solution. Therefore, we would like to ensure that if the bottleneck link is not the satellite link, CTCP does not adversely impact the performance of TCP flows not traversing the satellite.

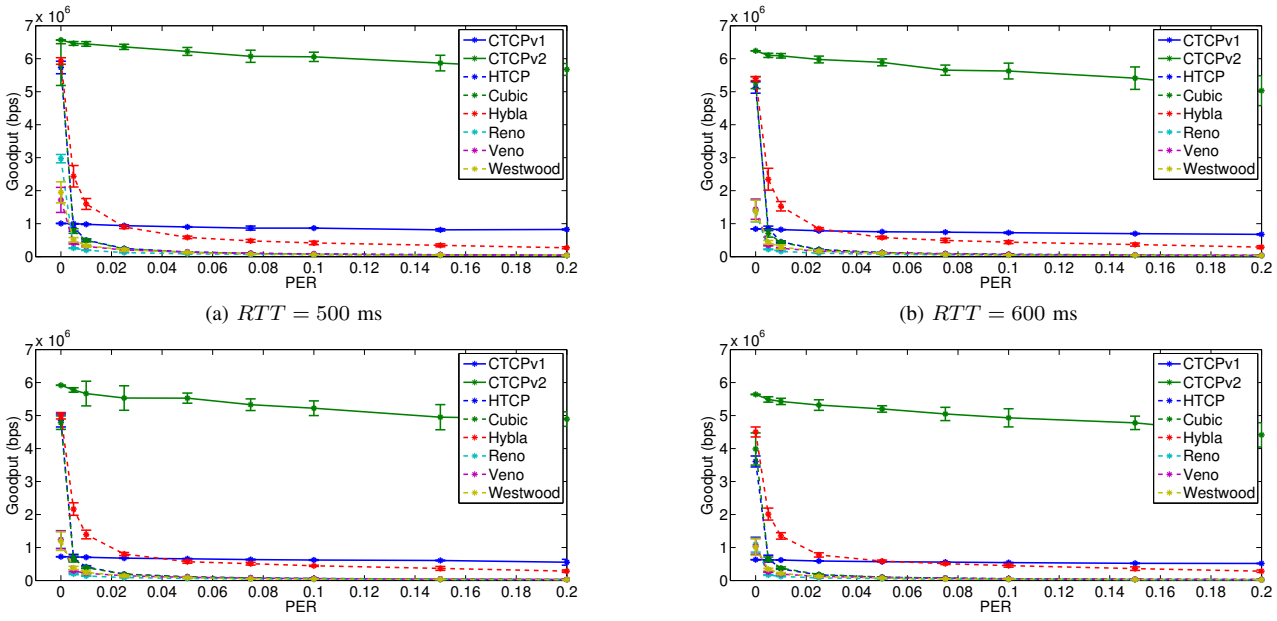
III. CTCP Performance in Satellite Networks

While previous results show that CTCP has great potential in networks with high packet losses and low RTT, no measurements exist for networks with large RTT. This section will explore the potential for CTCP to provide improved performance in environments with large delays using a testbed located at the Hamilton Institute, NUI Maynooth, Ireland.

The testbed used to collect measurements consists of commodity servers (Dell Poweredge 850, 3GHz Xeon, Intel 82571EB Gigabit NIC) connected via a router and gigabit switches. A diagram of the setup is shown in Figure 1. Sender and receiver machines used in the tests both run a Linux 2.6.32.27 kernel. The router is a commodity server running FreeBSD 4.11 and `ipfw-dummynet`. Data is transferred between the sender and receiver machines using `rsync` (version 3.0.4) and the appropriate TCP version.

Each version of TCP, other than CTCP, is implemented within the kernel making it easy to select the appropriate variant. In the case of CTCP, it is implemented in user space as a SOCKSv5 proxy with the forward proxy located on the client and the reverse proxy located on the server. Traffic between the proxies is sent using CTCP. Therefore, a client's request is first directed to the local forward proxy, transmitted to the reverse proxy, and then forwarded to the appropriate port on

 ISBN: 978-1-61208-317-9



(a) $RTT = 500$ ms (b) $RTT = 600$ ms

(c) $RTT = 700$ ms (d) $RTT = 800$ ms

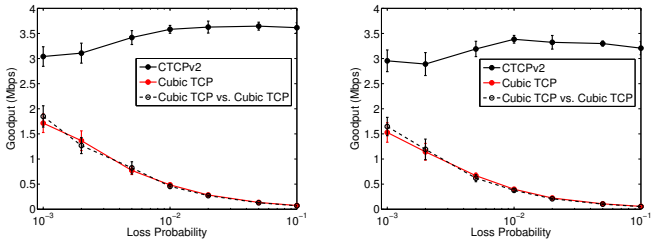
Figure 2: Comparison of TCP variants with varying PER and RTT with a link rate of 10 Mbps. Each bar shows the mean goodput, while the error bars show one standard deviation.

the server. The server responds using the reverse process. In order to use `rsync`, `proxychains` (version 3.1) was used to direct traffic to the proxy.

A series of tests were conducted using the following TCP variants: CTCP, Cubic, Hybla, Reno, Veno, and Westwood. A 20 MB file download is used, along with varying PER's and RTT's, to characterize the performance of each TCP version. Figure 2 provides a summary of each version's mean goodput as a function of the PER. Each test was run a minimum of three times and a maximum of ten times depending on the amount of time need to complete the 20 MB download.

The performance of CTCP, labeled "CTCP v1" in Figure 2, in networks with large RTT and low PER is significantly poorer than two of the TCP variants designed for these network types (i.e., TCP Hybla and TCP Cubic). For PER greater than 2.5%, CTCP begins to outperform both of these TCP variants for most of the RTTs tested. In fact, the goodput of CTCP remains relatively constant as the PER increases to 20% while the goodput of the other TCP variants approaches zero quickly.

The additive increase portion of CTCP's current congestion control algorithm is the primary reason for its poor performance at low PER. To overcome the challenges related to long RTT's, a modified version of CTCP, labeled "CTCP v2" in Figure 2, was implemented that increases $cwnd$ in a manner consistent with H-TCP (see [9] for more details). Because $cwnd$ is no longer dependent on the RTT, it can increase rapidly allowing it to use the available capacity more efficiently. In addition to the use of network coding and the unmodified multiplicative $cwnd$ back-off approach, CTCP is able to maintain a large throughput for PER's as high as 20%. In fact, measurements indicate that this modified version of CTCP provides a gain of approximately 21 times that of TCP



(a) $RTT = 500$ ms (b) $RTT = 800$ ms

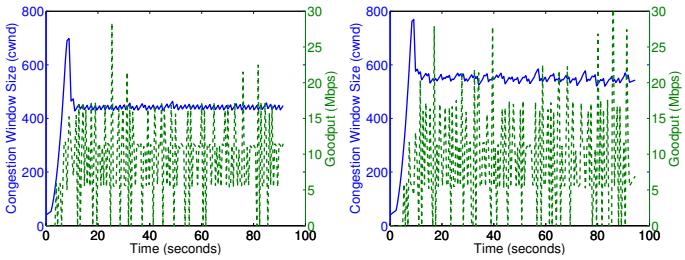
Figure 3: Goodput versus packet loss rate for (i) a Cubic TCP flow and CTCPv2 flow sharing a link (solid lines), and (ii) two Cubic TCP flows sharing a link (dashed line). The link rate is 5 Mbps and the error bars show one standard deviation.

Hybla for a PER of 20% and RTT of 500 ms over a link with a bandwidth of 10 Mbps.

In addition, preliminary testing has shown that this version of CTCP is friendly with existing TCP versions. Each subfigure in Figure 3 provides a comparison of the throughput for two tests. The dotted line labeled "Cubic TCP vs. Cubic TCP" shows the throughput obtained by one Cubic TCP flow competing against a second Cubic TCP flow. The solid lines show the second test where a Cubic TCP flow is competing against a CTCP flow. The indication of fairness is provided by the similarity of the solid Cubic TCP line and the dotted line.

While these results are promising, additional work is required. This is evident in the trace of the goodput and $cwnd$ shown in Figure 4. The instantaneous goodput is highly variable, which causes delay jitter as packets are delivered to higher layers. Possible causes of this may be an underestimate of the packet loss probability or an underestimate of the number of coded packets needed. Either case creates the distinct decode events shown in the figure. Regardless, the

 ISBN: 978-1-61208-317-9



(a) $PER = 0.5\%$ (b) $PER = 20\%$

Figure 4: Trace of CTCPv2 over a 10 Mbps link with a RTT of 500 ms. The dotted line shows the the instantaneous goodput and the solid line shows $cwnd$. The mean goodput in (a) and (b) is 9.19 Mbps and 8.92 Mbps respectively.

potential for greatly increasing performance at the transport layer is evident. Not only can throughput performance be drastically improved, but this solution appears, at first glance, to be backward compatible with existing TCP variants.

IV. Conclusions and Future Work

In Section II, an overview of CTCP is presented and a brief overview of CTCP's performance in networks with short RTT was provided. In Section III, CTCP's performance is compared with other TCP variants in networks with RTT's similar to those that would be observed with satellite communications. Measurements showed that the current implementation of CTCP performed worse than some existing TCP versions for small PER, but outperformed other TCP versions for high PER. One of the primary causes of this was discussed and an alternate $cwnd$ increase mechanism is used to highlight that some minor changes to the current congestion control algorithm can significantly improve performance in these environments. While previous sections discussed several areas of future research such as this, the remainder of this section will introduce additional future research directions.

Additional research into CTCP's congestion control algorithm is required. First, the current implementation relies on the RTT of the path in order to increase $cwnd$. For connections with large RTT's, this is obviously an issue. More aggressive methods for increasing $cwnd$, such as using a H-TCP like mechanism, need to be thoroughly developed while still maintaining interoperability with network coding and fairness with legacy TCP variants. Second, the method for determining whether a packet is lost due to congestion or due to a poor link is also an issue. The multiplicative back off method currently used (i.e., $\beta = {RTT_{\min}}/{RTT}$) works well when jitter in the RTT measurement is primarily caused by the filling of queues. It fails to work properly when the delay jitter is caused by something else. For example, if the delay jitter is caused by a particular medium-access (MAC) method, it is likely that $\beta < 1$ causing $cwnd$ to collapse. This was observed in measurements taken over a WiMax network [3] where the MAC's scheduling algorithm caused large variations in the RTT. Methods using feedback from the network, such as LT-TCP, which uses explicit congestion notification (ECN) [7], are possibilities although we would like to ensure that CTCP operates irrespective of lower layer implementations.

The use of network coding in CTCP is critical for overcoming packet losses and providing high throughput, but little is understood about how to adjust the number of packets coded together (i.e., the coding window). The current implementation of CTCP uses a fixed size block, or generation, scheme for generating network coded packets. This is not optimal given the user's requirements since there is an inherent tradeoff between throughput and delay as the block, or generation, size is changed [12]. Dynamically adjusting the block size, or using a sliding coding window approach, to meet the user's throughput/delay requirements is also a topic of ongoing research. Furthermore, the interaction between congestion avoidance and network coding is not fully understood. The current implementation treats both congestion avoidance and network coding separately, yet there is evidence that intelligently merging the two can provide a performance increase.

In summary, CTCP has potential to greatly improve network performance over existing transport layer protocols in the presence of both high packet error rates and round-trip times. Initial measurements have shown significant gains in goodput over existing TCP versions, but additional research is needed to tune both the congestion control algorithm and network coding parameters to ensure proper functionality.

Acknowledgment

This work is sponsored, in part, by the Assistant Secretary of Defense (ASD R&E) under Air Force Contract # FA8721-05-C-0002. Opinions, interpretations, recommendations and conclusions are those of the authors and are not necessarily endorsed by the United States Government.

References

[1] C. Caini, R. Firrincieli, M. Marchese, T. d. Cola, M. Luglio, C. Roseti, N. Celandroni, and F. Potorti, "Transport Layer Protocols and Architectures for Satellite Networks," International Journal of Satellite Communications and Networking, vol. 25, no. 1, 2007, pp. 1–26.

[2] A. Pirovano and F. Garcia, "A New Survey on Improving TCP Performances Over Geostationary Satellite Link," Network and Communication Technologies, vol. 2, no. 1, Jan. 2013.

[3] M. Kim, J. Cloud, A. ParandehGheibi, L. Urbina, K. Fouli, D. Leith, and M. Médard, "Network Coded TCP (CTCP)," CoRR, vol. abs/1212.2291v2, 2012. [Online]. Available: http://arxiv.org/abs/1212.2291

[4] S. Ha, I. Rhee, and L. Xu, "CUBIC: A New TCP-Friendly High-Speed TCP Variant," SIGOPS Oper. Syst. Rev., vol. 42, no. 5, Jul. 2008, pp. 64–74.

[5] C. Caini and R. Firrincieli, "TCP Hybla: A TCP Enhancement for Heterogeneous Networks," International Journal of Satellite Communications and Networking, vol. 22, no. 5, 2004, pp. 547–566.

[6] S. Trivedi, S. Jaiswal, R. Kumar, and S. Rao, "Comparative Performance Evaluation of TCP Hybla and TCP Cubic for Satellite Communication Under Low Error Conditions," in IMSAA, 2010, pp. 1–5.

[7] B. Ganguly, B. Holzbauer, K. Kar, and K. Battle, "Loss-Tolerant TCP (LT-TCP): Implementation and Experimental Evaluation," in MILCOM, 2012, pp. 1–6.

[8] V. Sharma, S. Kalyanaraman, K. Kar, K. K. Ramakrishnan, and V. Subramanian, "MPLOT: A Transport Protocol Exploiting Multipath Diversity Using Erasure Codes, year = 2008," in INFOCOM, pp. 121–125.

[9] D. Leith and R. Shorten, "H-TCP: TCP for High-Speed and Long-Distance Networks," 2004.

[10] T. Ho, M. Médard, R. Koetter, D. Karger, M. Effros, J. Shi, and B. Leong, "A Random Linear Network Coding Approach to Multicast," IEEE Transactions on Information Theory, vol. 52, no. 10, 2006, pp. 4413–4430.

[11] A. Urke, L. Braten, and K. Ovsthus, "TCP Challenges in Hybrid Military Satellite Networks; Measurements and Comparison," in MILCOM, 2012, pp. 1–6.

[12] W. Zeng, C. Ng, and M. Médard, "Joint Coding and Scheduling Optimization in Wireless Systems with Varying Delay Sensitivities," in SECON, 2012, pp. 416–424.

 ISBN: 978-1-61208-317-9