Stochastic Modeling for Self-evolving Botnets in Infection Control Environments

Koki Hongyo*, Tomotaka Kimura†, Takanori Kudo‡, Yoshiaki Inoue§, and Kouji Hirata*
* Faculty of Engineering, Kansai University, Osaka 564-8680, Japan, Email: {k896955, hirata}@kansai-u.ac.jp
† Faculty of Science and Engineering, Doshisha University, Kyoto 610-0321, Japan, Email: tomkimur@mail.doshisha.ac.jp
‡ Faculty of Science and Engineering, Setsunan University, Osaka 572-8508, Japan, Email: t-kudo@ele.setsunan.ac.jp
§ Graduate School of Engineering, Osaka University, Osaka 565-0871, Japan, Email: yoshiaki@comm.eng.osaka-u.ac.jp

***Abstract*—The concept of self-evolving botnets, where computing resources of infected hosts are exploited to discover unknown vulnerabilities and the botnets evolve autonomously, has been introduced and their threats have been shown in the literature. In order to protect networks from the self-evolving botnets, this paper provides an epidemic model taking into account the infection routes in infection control environments to which countermeasures against the self-evolving botnets are applied. We show the behaviors of the epidemic model through simulation experiments.**

***Keywords*–Botnet; machine learning; epidemic model; continuous-time Markov chain.**

I. Introduction

Recently, machine learning techniques have been widely used and achieved significant results in various research areas. In addition, some researchers have proposed vulnerability discovery methods that discover bugs and vulnerabilities with machine learning techniques [5][6]. Although the main purpose of these methods is to protect software, they can be used for discovering unknown security holes and exploited for illegal attacks by malicious attackers. To perform illegal attacks, malicious attackers often control botnets, which consist of many infected hosts named zombie computers. The malicious attackers can discover unknown vulnerabilities with distributed machine learning using the computing resources of the zombie computers.

Based on these facts, in [4], Kudo et al. have introduced a new concept named self-evolving botnets. Self-evolving botnets discover vulnerabilities by performing distributed machine learning with computing resources of zombie computers and evolve autonomously based on the vulnerabilities. Accordingly, they infect other hosts and make themselves bigger. The authors have shown that the infectivity of self-evolving botnets is very high, compared with conventional botnets. In response, in [2], Hongyo et al. have proposed some epidemic models that consider countermeasures against self-evolving botnets and shown their effectiveness.

In this paper, we propose an epidemic model for self-evolving botnets taking into account the infection routes of the botnets in infection control environments to which some countermeasure methods are applied. Because the infectivity of the botnets often depends on infection routes, the proposed epidemic model expresses the infection routes by overlay networks that are constructed according to relationships among hosts. The proposed epidemic model makes continuous-time Markov chains with the overlay networks and show the behavior of the self-evolving botnets in infection control environments. The rest of this paper is organized as follows. Section II explains our proposed epidemic model. We then evaluate it in Section III.

II. Epidemic model for self-evolving botnets

We use an Susceptible-Infected-Recovered-Susceptible (SIRS) model to represent the state of each host in a network. In the SIRS model, “S” means that the host has vulnerabilities, “I” means that the host is infected, and “R” means that the host has no known vulnerabilities. Each host belongs to one of the states. We assume that hosts in the state R can get infected by unknown vulnerabilities which are discovered by a self-evolving botnet. Hosts in the state S move to the state I when they get infected by attacks of a botnet. Then, the hosts are embedded in the botnet. Hosts in the state S and the state I move to the state R when known vulnerabilities and the botnet malware, respectively, are removed from the hosts by suitable means, such as OS updates and anti-virus software. Note that we assume that all vulnerabilities are simultaneously removed in these cases. When the botnet discovers a new vulnerability, all hosts in the state R move to the state S because the botnet can infect the hosts by using the discovered vulnerability.

The proposed epidemic model considers relationships among hosts in the above SIRS model because infection routes of the self-evolving botnets depend on the relationships, e.g., their friendships, frequently accessed web sites, and physical network environments. To express the relationships, we use an overlay network consisting of hosts. Hosts in state I can infect only adjacent susceptible hosts on the overlay network. Under this assumption, the proposed epidemic model formulates the infection process of the self-evolving botnet as a continuous-time Markov chain, where the occurrence of each event a)-d) described below in the SIRS model follows a Poisson process.

(a) A new vulnerability is discovered by the self-evolving botnet according to a Poisson process with the discovery rate $\eta(v+1)$, where v denotes the number of infected hosts and η denotes the discovery rate of a new vulnerability by each infected host. The discovery rate is proportional to the number of infected hosts, which means that the self-evolving botnet performs distributed machine learning with the computing resources of the infected hosts. When this event occurs, all the hosts in the state R moves to the state S.

(b) Each host in the state S removes its own vulnerabilities according to a Poisson process with the recovery rate δ_S, and then moves to the state R.

(c) Each host in the state I infects an adjacent host in the state S on the overlay network according to a Poisson process with the infection rate α. In this case, the adjacent host moves to the state I.

(d) Each host in the state I removes the botnet malware according to a Poisson process with the removal rate δ_{I}, and then moves to the state R.

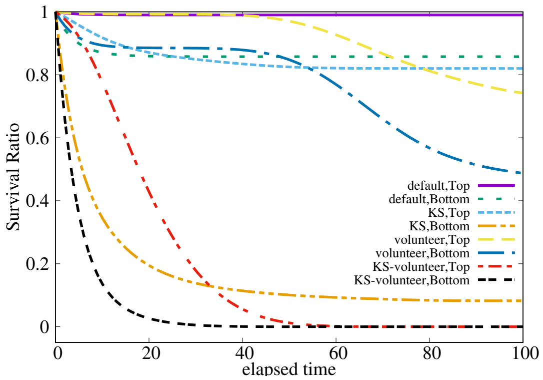

Figure 1. Botnet survival ratio.

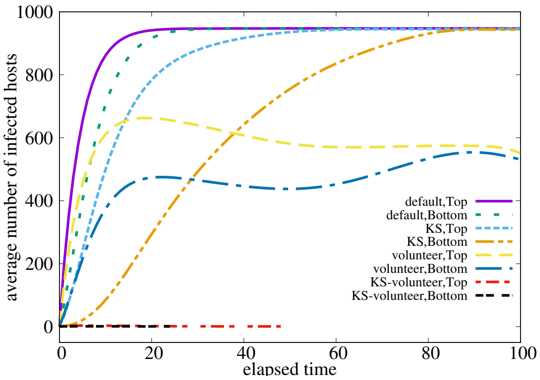

Figure 2. Average number of infected hosts.

We then consider countermeasures against the self-evolving botnet. As the countermeasures, we adopt a Kill-Signal (KS) model and a volunteer model. The KS model proposed in [3] uses a warning signal called Kill-Signal having information on known vulnerabilities. In the KS model, the hosts in the state R send a Kill-Signal to susceptible hosts. The hosts receiving the Kill-Signal can know their vulnerabilities due to the Kill-Signal and repair them. The volunteer model aims to discover and repair unknown vulnerabilities with computing resources of volunteer hosts before the self-evolving botnet discover the vulnerabilities, so that it suppresses the evolution of the self-evolving botnet. In this paper, for simplicity, we assume that all uninfected hosts (i.e., hosts in the states S or R) belong to a volunteer group and a network administrator can use their computing resources to discover vulnerabilities. The information on discovered vulnerabilities are shared by all uninfected hosts. These models add or replace the events in the Markov chain as follows.

(e) Each host in the state R sends a Kill-Signal to an adjacent susceptible host on the overlay network according to a Poisson process with the sending rate β_{S}. In this case, the adjacent host moves to the state R.

(f) A new vulnerability is discovered by the volunteer group according to a Poisson process. Accordingly, the infectivity of the self-evolving botnet is weakened. To represent this behavior, the volunteer model replaces the discovery rate of the self-evolving botnet described in event (a) with $\eta(v+1)/(\sigma(N-v)+1)$, where σ denotes the vulnerability discovery rate of a volunteer host.

III. Evaluation

To examine the behavior of the proposed epidemic model, we conduct simulation experiments. We assume that there are $N = 1,000$ hosts in a network and the overlay network is constructed based on the Barabasi-Albert model [1], where the average degree of hosts is 20. One host is infected and all the other hosts are in the susceptible state at time $t = 0$. We refer to the infected host at time $t = 0$ as the initial infected host. The parameters are set to be $\eta = 0.05$, $\delta_{\mathrm{S}} = \beta_{\mathrm{S}} = 0.1$, $\delta_{\mathrm{I}} = 0.1$, $\alpha = 0.1$, and $\sigma = 0.3$.

Figure 1 shows the botnet survival ratio as a function of the elapsed time. The botnet survival ratio means the ratio of the number of samples in which one or more infected hosts still exist at time t to the total number of samples. In this figure, "Top" (resp. "Bottom") indicates the result in the case where a host with the maximum (resp. minimum) closeness centrality is selected as the initial infected host. Furthermore, "default" represents the result of the self-evolving botnets without countermeasures, "KS" represents the result of the KS model, "volunteer" represents the result of the volunteer model, and "KS-volunteer" represents the result of the mixed model of the KS and volunteer models. As we can see from this figure, the botnet survival ratio is large when selecting a host with the maximum closeness centrality as the initial infected host. We also observe that the botnet survival ratio of "default" is very high. "KS" decreases the botnet survival ratio at early stage, but does not decrease with the time elapsed. On the other hand, "volunteer" first does not decrease the botnet survival ratio, but gradually decreases it. This result implies that the volunteer model can weaken the capability of the self-evolving botnet even though the self-evolving botnet spreads. Furthermore, "KS-volunteer" eliminates the self-evolving botnet early in all samples.

Figure 2 shows the average number of infected hosts of samples in which there exist infected hosts at time t as a function of the elapsed time t. As shown in this figure, the average numbers of infected hosts of "default" and "KS" increase and converge to a high value, regardless of the closeness centrality of the initial infected hosts. On the other hand, "volunteer" can reduce the the average number of infected hosts constantly. Furthermore, "KS-volunteer" can eliminate completely the self-evolving botnet.

References

[1] A. Barabasi and R. Albert, "Emergence of scaling in random networks," *Science*, vol. 286, pp.509–512, 1999.

[2] K. Hongyo et al., "Modeling of countermeasure against self-evolving botnets," in *Proc. ICCE-TW 2017*, Taipei, Taiwan, Jun. 2017, pp. 1–2.

[3] J. O. Kephart and S. R. White, "Measuring and modeling computer virus prevalence," in *Proc. the 1993 IEEE Computer Society Symposium on Research in Security and Privacy*, Oakland, CA, May. 1993.

[4] T. Kudo et al., "Behavior analysis of self-evolving botnets," in *Proc. CITS 2016*, Kunming, China, Jul. 2016.

[5] R. Scandariato, J. Walden, A. Hovsepyan, and W. Joosen, "Predicting vulnerable software components via text mining," *IEEE Transactions on Software Engineering*, vol. 40, no. 10, pp. 993–1006, 2014.

[6] F. Yamaguchi, F. Lindner, and K. Rieck, "Vulnerability extrapolation: assisted discovery of vulnerabilities using machine learning," in *Proc. USENIX conference on Offensive Technologies*, San Francisco, CA, Aug. 2011.