Evaluation of countermeasure against future malware evolution with deterministic modeling

著者	Shimizu Koki, Kumai Yuya, Motonaka Kimiko, Kimura Tomotaka, Hirata Kouji
会議概要（会議名，開催地，会期，主催者等）	Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC 2019), November 18-21, 2019, Lanzhou, China
URL	http://hdl.handle.net/10112/00017626

Evaluation of countermeasure against future malware evolution with deterministic modeling

Koki Shimizu*, Yuya Kumai*, Kimiko Motonaka†, Tomotaka Kimura‡, and Kouji Hirata†
* Graduate School of Science and Engineering, Kansai University, Osaka, Japan
E-mail: {k388626, k944445}@kansai-u.ac.jp
† Faculty of Engineering Science, Kansai University, Osaka, Japan
E-mail: {motonaka, hirata}@kansai-u.ac.jp
‡ Faculty of Science and Engineering, Doshisha University, Kyoto, Japan
E-mail: tomkimur@mail.doshisha.ac.jp

***Abstract*—Recently, machine learning technologies have dramatically evolved. Accordingly, the concept of self-evolving botnets has been introduced, which discover vulnerabilities of hosts by distributed machine learning using the computational resources of infected hosts, and infect other hosts by attacks using the discovered vulnerabilities. The infectability of the self-evolving botnets is too strong compared with conventional botnets, so that such new botnets will become the serious threat to future network society including 5G and IoT environments. In this paper, we consider a volunteer model that discovers unknown vulnerabilities earlier than self-evolving botnets by distributed computing using volunteer hosts' resources and repairs the vulnerabilities. We propose deterministic modeling for the volunteer model. Through numerical calculations, we evaluate the performance of the volunteer model against self-evolving botnets.**

I. Introduction

Recently, machine learning techniques have been widely used in many research fields and achieved significant results because of the recent dramatic evolution of deep learning [1], [2]. The literature has proposed vulnerability discovery methods that discover bugs and vulnerabilities with static code analysis and machine learning techniques [7], [8]. Although the main purpose of these methods is to protect software, they can be used for discovering unknown vulnerabilities and exploited for illegal attacks by malicious attackers. To perform illegal attacks such as DDoS attacks, malicious attackers often construct a botnet, which consists of hosts getting infected with the botnet malware [6]. In the past, botnets that consist of more than a million zombie computers have appeared.

Based on these backgrounds, the authors in [4] have introduced a new concept named self-evolving botnets. The self-evolving botnets discover vulnerabilities of hosts by performing distributed machine learning with computing resources of infected hosts (i.e., zombie computers). Accordingly, they infect the hosts by attacks using the discovered vulnerabilities and make themselves bigger by absorbing the hosts. The authors in [4] have proposed a stochastic epidemic model of the self-evolving botnets. The epidemic model represents the infection dynamics of the self-evolving botnets with a continuous-time Markov chain. Furthermore, in [5], the authors have introduced a deterministic epidemic model of the self-evolving botnets, which represents their infection dynamics by ordinary differential equations. They have shown that the infectability of the self-evolving botnets is much stronger than conventional botnets, through numerical experiments.

The appearance of such new botnets will become the serious threat to future network society including 5G and IoT environments. In order to counter the threat of the self-evolving botnets, in this paper, we introduce a countermeasure model named volunteer model against infection and spread of the botnet malware. The volunteer model extends the basic concept discussed in [3] and aims at countering the self-evolving botnets by discovering unknown vulnerabilities earlier than the self-evolving botnets. In the volunteer model, volunteer hosts' computing resources are used to discover the unknown vulnerabilities in non-infected hosts. Those discovered vulnerabilities are repaired, and thus the hosts can be protected from the infection of self-evolving botnets. Furthermore, we propose deterministic modeling for the volunteer model. In the deterministic epidemic model, the infection dynamics of self-evolving botnets under the situation where the volunteer model is applied are represented by ordinary differential equations. Through numerical calculations, we evaluate the performance of the volunteer model against self-evolving botnets.

The rest of this paper is organized as follows. Section II discusses the epidemic model for self-evolving botnets. In Section III, we explain the volunteer model. In Section IV, we discuss the behavior of the volunteer model with the results of numerical calculations. We conclude the paper in Section V.

II. Deterministic epidemic model of self-evolving botnets [5]

A. SIRS model

We first explain the deterministic epidemic model of self-evolving botnets in situations where the volunteer model is not applied. This deterministic epidemic model, which represents the system states (i.e., the infection dynamics of the self-evolving botnets), works based on the SIRS (Susceptible-Infected-Recovered-Susceptible) model, which expresses states of each host.

The SIRS model consists of three states Susceptible (S), Infected (I), and Recovered (R) as shown in Fig. 1. The

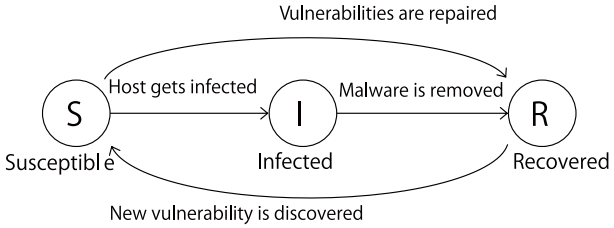

Fig. 1. SIRS model.

meaning of each state is as follows:

S : The host has at least one known vulnerabilities.
I : The host is infected with the botnet malware.
R : The host has no known vulnerabilities.

We assume that hosts belonging to the state R do not get infected by known vulnerabilities. However, when the self-evolving botnet discovers a new vulnerability, their states transition to the state S. In this case, they can get infected by the new vulnerability.

In the SIRS model, the transition from the susceptible state S to the infected state I means that a susceptible host gets infected by attack of the self-evolving botnet. In this case, the infected host is taken into the self-evolving botnet and its computing resource is used to discover new vulnerabilities. The transition from the susceptible state S to the recovered state R means that a susceptible host removes known vulnerabilities from itself. Also, the transition from the infected state I to the recovered state R indicates that an infected host removes the botnet malware from itself, where we assume that all known vulnerabilities are removed at the same time. When the self-evolving botnet discovers a new vulnerability, all hosts belonging to the state R transitions to the state S. This transition means that the botnet becomes able to infect the host by using the discovered vulnerability.

In [5], the authors have introduced a deterministic epidemic model of self-evolving botnets based on the SIRS model. In the deterministic epidemic model, the infection dynamics of self-evolving botnets is represented by ordinary differential equations.

B. Epidemic model

The deterministic epidemic model of self-evolving botnet is based on the SIRS model. Specifically, it is defined as follows. Let $S(t)$, $I(t)$, and $R(t)$ denote the numbers of hosts in the states S, I, and R, respectively, at time t. We assume that the total number N of hosts is fixed and it does not change over time t (i.e., $N = S(t) + I(t) + R(t)$). The change rate of the number of hosts in each state is given by the following ordinary differential equations:

$$\frac{dS(t)}{dt} = -\alpha S(t)I(t) + \gamma I(t)R(t) - \delta S(t), \quad (1)$$

$$\frac{dI(t)}{dt} = \alpha S(t)I(t) - \beta I(t), \quad (2)$$

$$\frac{dR(t)}{dt} = \beta I(t) - \gamma I(t)R(t) + \delta S(t), \quad (3)$$

where α denotes the malware infection rate per infected host, and thus $\alpha S(t)I(t)$ indicates the average number of susceptible hosts getting infected per unit time at time t in (1) and (2). β denotes the malware elimination rate per infected host. Therefore, $\beta I(t)$ indicates the average number of infected hosts eliminating the botnet malware from themselves per unit time at time t in (2) and (3). γ denotes the new vulnerability discovery rate per infected host in the self-evolving botnet. In (1) and (3), $\gamma I(t)R(t)$ indicates the average number of recovered hosts whose vulnerability is discovered by the self-evolving botnet per unit time at time t, assuming that vulnerability discovery is performed by infected hosts. δ denotes the repair rate per susceptible host, and thus $\delta S(t)$ in (1) and (3) means the average number of susceptible hosts repairing their own vulnerabilities per unit time at time t.

III. Deterministic epidemic modeling for the volunteer model

A. Assumption

Because the self-evolving botnet has very high infectability, it is very difficult for each host to individually protect itself from the infection. To overcome this difficulty, the volunteer model counters the self-evolving botnet, using the computing resources of volunteer hosts to discover new vulnerabilities and protect the hosts before the self-evolving botnet discovers them. In this paper, we represent the infection dynamics of the volunteer model under the following assumptions.

1) There exists one volunteer group consisting of all volunteer hosts in a given network.
2) Each susceptible host and each recovered host can join the volunteer group (i.e., becoming a volunteer host). The probability that a host becomes a volunteer host is proportional to the vulnerability discovery capability of the volunteer hosts. This is because we assume that the participation of new hosts to the volunteer group is encouraged when the vulnerability protection becomes effective with the increase in the vulnerability discovery capability.
3) The volunteer group provides the information on vulnerability discovery to all hosts in the network, and thus they can repair the vulnerability.
4) Hosts belong to the volunteer group share the information on their known vulnerabilities each other.
5) Each volunteer host can freely leave the volunteer group.

Fig. 2 represents the state transition diagram of each host in the volunteer model, which follows these assumptions and is based on the SIRS model shown in Fig. 1. The meaning of each state is as follows:

S_1 : The host belongs to the susceptible state but not the volunteer group.
S_2 : The host belongs to the susceptible state and the volunteer group.
I : The host belongs to the infected state.
R_1 : The host belongs to the recovered state but not the volunteer group.

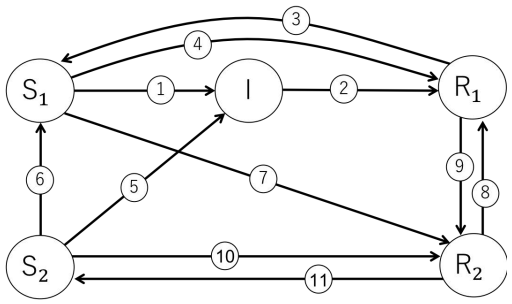

Fig. 2. State transitions of each host in the volunteer model.

R_2 : The host belongs to the recovered state and the volunteer group.

In the volunteer model, the susceptible state S and the recovered state R are divided into two states "S_1, S_2" and "R_1, R_2", respectively. S_1 and R_1 indicates that the host does not belong to the volunteer group. On the other hand, S_2 and R_2 indicates the host belongs to the volunteer group. Each state transition ①-⑪ in Fig. 2 means the occurrence of the following event:

a) The host gets infected by an attack of an infected host (①, ⑤).
b) The host eliminates the botnet malware from itself (②).
c) The host repairs known vulnerabilities from itself (④, ⑩).
d) The host leaves the volunteer group (⑥, ⑧).
e) The host joins the volunteer group (⑦, ⑨).
f) The self-evolving botnet discovers a new vulnerability of the host (③, ⑪).

In the event e), susceptible hosts transition to the recovered state R_2 immediately after they join the volunteer group because we assume that hosts belonging to the volunteer group share the information on vulnerabilities known by the volunteer group.

B. Epidemic model

We assume that there exist N hosts in a network. Let $S_1(t)$ and $S_2(t)$ denote the numbers of hosts belonging to the susceptible states S_1 and S_2, respectively, at time t. Let $I(t)$ denote the number of infected hosts, and $R_1(t)$ and $R_2(t)$ denote the numbers of hosts belonging to the recovered states R_1 and R_2, respectively at time t. The infection dynamics of the volunteer model is represented by the following ordinary differential equations, where $N = S_1(t) + S_2(t) + I(t) + R_1(t) + R_2(t)$:

$$\begin{aligned}\frac{dS_1(t)}{dt} &= -\alpha_1 S_1(t)I(t) + \theta S_2(t) - \delta_1 S_1(t) \\ &\quad - \epsilon S_1(t)(S_2(t) + R_2(t) + 1) \\ &\quad + \gamma I(t) R_1(t) \frac{\gamma I(t)}{\psi(S_2(t) + R_2(t)) + \gamma I(t)},\end{aligned} \tag{4}$$

$$\begin{aligned}\frac{dS_2(t)}{dt} &= -\alpha_2 S_2(t)I(t) - \theta S_2(t) - \delta_2 S_2(t)R_2(t) \\ &\quad + \gamma I(t) R_2(t) \frac{\gamma I(t)}{\psi(S_2(t) + R_2(t)) + \gamma I(t)},\end{aligned} \tag{5}$$

$$\frac{dI(t)}{dt} = \alpha_1 S_1(t)I(t) + \alpha_2 S_2(t)I(t) - \beta I(t), \tag{6}$$

$$\begin{aligned}\frac{dR_1(t)}{dt} &= \delta_1 S_1(t) - \mu R_1(t)(S_2(t) + R_2(t) + 1) + \beta I(t) \\ &\quad + \lambda R_2(t) - \gamma I(t) R_1(t) \frac{\gamma I(t)}{\psi(S_2(t) + R_2(t)) + \gamma I(t)},\end{aligned} \tag{7}$$

$$\begin{aligned}\frac{dR_2(t)}{dt} &= \delta_2 S_2(t)R_2(t) + \epsilon S_1(t)(S_2(t) + R_2(t) + 1) \\ &\quad + \mu R_1(t)(S_2(t) + R_2(t) + 1) - \lambda R_2(t) \\ &\quad - \gamma I(t) R_2(t) \frac{\gamma I(t)}{\psi(S_2(t) + R_2(t)) + \gamma I(t)},\end{aligned} \tag{8}$$

where α_1 and α_2 denote the malware infection rates for susceptible hosts in S_1 and S_2, respectively, per infected host. β denotes the malware elimination rate per infected host. δ_1 and δ_2 denote the vulnerability repair rates per susceptible hosts in S_1 and S_2, respectively. ϵ and μ denote the rates of joining the volunteer group per susceptible host and recovered host, respectively. θ and λ denote the rates of leaving the volunteer group per susceptible host and recovered host, respectively. γ and ψ denote the new vulnerability discovery rates per host in the self-evolving botnet and the volunteer group, respectively.

In (4), (5), and (6), $\alpha_1 S_1(t)I(t)$ and $\alpha_2 S_2(t)I(t)$ represent the average numbers of hosts in the susceptible states S_1 and S_2, respectively, that get infected per unit time at time t. They correspond to the state transitions ① and ⑤. In (6) and (7), $\beta I(t)$ represents the average number of infected hosts that eliminate the botnet malware from themselves per unit time at time t, which correspond to the state transition ②. $\delta_1 S_1(t)$ and $\delta_2 S_2(t)R_2(t)$ in (4), (5), (7), and (8) mean the average numbers of hosts in the susceptible states S_1 and S_2, respectively, that repair their own vulnerabilities per unit time at time t. Because we assume that recovered hosts in the volunteer group give susceptible volunteer hosts the information on known vulnerabilities, we represent the average number with $\delta_2 S_2(t)R_2(t)$ for the volunteer hosts. They correspond to the state transitions ④ and ⑩.

In (4), (7), and (8), $\epsilon S_1(t)(S_2(t) + R_2(t) + 1)$ and $\mu R_1(t)(S_2(t) + R_2(t) + 1)$ represent the average numbers of susceptible hosts and recovered hosts, respectively, joining the volunteer group per unit time at time t, which correspond to the state transitions ⑦ and ⑨. The vulnerability discovery capability increases with the number of hosts in the volunteer group, so that the participation of new hosts to the volunteer group is encouraged. On the other hand, $\theta S_2(t)$ and $\lambda R_2(t)$ in (4), (5), (7), and (8) means the average numbers of susceptible hosts and recovered hosts, respectively, leaving the volunteer group per unit time at time t. They correspond to the state transitions ⑥ and ⑧.

In (4), (5), (7), and (8), $\gamma I(t)R_1(t)$ and $\gamma I(t)R_2(t)$ indicate the average numbers of recovered hosts in R_1 and R_2, respectively, whose vulnerability is discovered by the self-evolving botnet per unit time at time t. We assume that they are weakened according to the discovery capability of the volunteer group. Specifically, they are multiplied by the

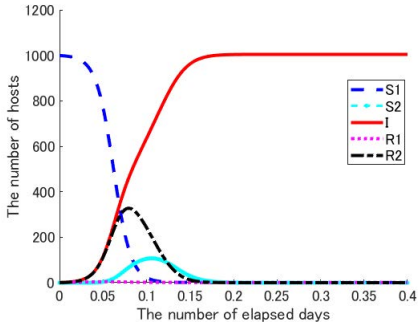

Fig. 3. The number of hosts in each state [$t = 0$ to 0.4] ($\alpha_1 = \alpha_2 = 0.1$, $\epsilon = \mu = 0.1$).

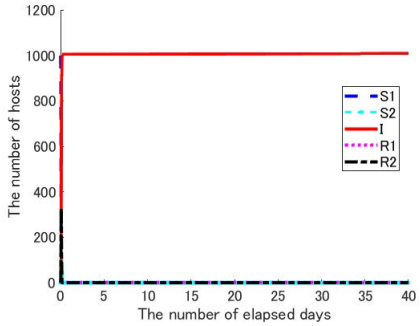

Fig. 4. The number of hosts in each state [$t = 0$ to 40] ($\alpha_1 = \alpha_2 = 0.1$, $\epsilon = \mu = 0.1$).

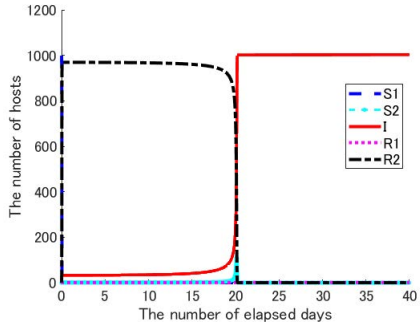

Fig. 5. The number of hosts in each state ($\alpha_1 = \alpha_2 = 0.1$, $\epsilon = \mu = 0.197$).

ratio of $\gamma I(t)$ to $\psi(S_2(t) + R_2(t)) + \gamma I(t)$, where $\gamma I(t)$ means the discovery capability of the self-evolving botnets and $\psi(S_2(t) + R_2(t))$ means the discovery capability of the volunteer group.

IV. Numerical Calculations

A. Model

In this paper, we examine the infection dynamics of the volunteer model through numerical calculations. The total number $N$of hosts in a network is equal to 1,000. The initial state of the system is assumed to be $(S_1(0), S_2(0), I(0), R_1(0), R_2(0)) = (999, 0, 1, 0, 0)$. Specifically, there is one infected host and the other hosts have vulnerabilities, which do not belong to the volunteer group. The system parameters in (4)-(8) are set to be $\beta = 0.1$, $\gamma = 0.1$, $\delta_1 = \delta_2 = 0.1$, $\psi = 0.1$, and $\theta = \lambda = 0.1$.

B. Results

Figs. 3 and 4 show the number of hosts in each state as a function of elapsed days, where the malware infection rates α_1 and α_2 are set to be 0.1 and the rates ϵ and μ of joining the volunteer group are set to be 0.1. Note that these figures use different time scales, i.e., [0-0.4] and [0-40]. From these figures, we observe that the number of recovered hosts belonging to the volunteer group first increases, but it immediately decreases and becomes almost 0. We also observe that the number of susceptible hosts decreases at an early stage and the number of infected hosts rapidly increases, which means that almost all susceptible hosts get infected with the botnet malware. The vulnerability discovery capability of the self-evolving botnet becomes strong with the increase in the number of infected hosts. Therefore, even if infected hosts eliminate the botnet malware from themselves and transition to the recovered state, they are immediately discovered a new vulnerability by the self-evolving botnet. Therefore, they easily transition to the susceptible state, and then get infected with the botnet malware. Therefore, the volunteer group does not work well in this case.

Here, we examine the impact of the system parameter ϵ and μ, which are the rates of joining the volunteer group per susceptible host and recovered host, respectively. Fig. 5 shows the number of hosts in each state as a function of elapsed days, where $\alpha_1 = \alpha_2 = 0.1$ and $\epsilon = \mu = 0.197$. As we can see from this figure, the number of volunteer hosts keeps high value. However, after about 20 days, the number of volunteer hosts immediately decreases and the number of infected hosts increases accordingly. This result indicates that when the number of volunteer hosts falls below a certain value, the volunteer model cannot prevent the spread of the self-evolving botnet. Figs. 6 and 7 show the number of hosts in each state as a function of elapsed days with different time scales, where $\alpha_1 = \alpha_2 = 0.1$ and $\epsilon = \mu = 0.198$. From these figures, we observe that the number of volunteer hosts does not decrease and almost all the hosts become the volunteer hosts. In this case, the volunteer model can prevent the spread of the self-evolving botnet. The effect of the volunteer model is very sensitive to the system parameters ϵ and μ.

We then examine the impact of the system parameter α_1 and α_2, which are the malware infection rates for hosts in S_1 and S_2, respectively. Figs. 8 and 9 show the number of hosts in each state as a function of elapsed days with different time scales, where $\alpha_1 = \alpha_2 = 0.2$ and $\epsilon = \mu = 0.1$. As we can see from these figures, the spreading speed of the self-evolving botnet is faster than the case of $\alpha_1 = \alpha_2 = 0.1$ shown in Fig. 3. This is because the malware infection rate is higher. Figs. 10 shows the number of hosts in each state as a function of elapsed days, where $\alpha_1 = \alpha_2 = 0.2$ and $\epsilon = \mu = 0.43$. Also, Figs. 11 shows the number of hosts in each state as a function of elapsed days, where $\alpha_1 = \alpha_2 = 0.2$ and $\epsilon = \mu = 0.44$. Similar to the results in Figs. 5-7, we observe that the number of volunteer hosts does not decrease and almost all the hosts become the volunteer hosts when the values of ϵ and μ are higher than a certain level.

V. Conclusion

This paper introduced a volunteer model to countermeasure self-evolving botnets. Through numerical calculations, we showed that the volunteer model efficiently works when many hosts join the volunteer groups. As future work, we will consider how to encourage hosts to join the volunteer model. This paper assumes that hosts join the volunteer group according to the number of volunteer hosts. This is because the effect of vulnerability discovery and protection increases with

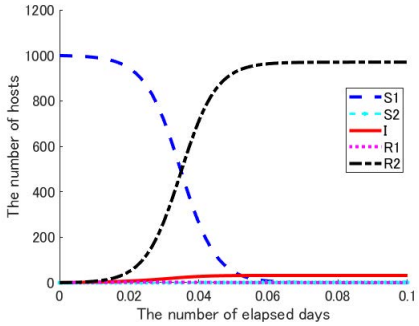

Fig. 6. The number of hosts in each state [$t = 0$ to 0.1] ($\alpha_1 = \alpha_2 = 0.1$, $\epsilon = \mu = 0.198$).

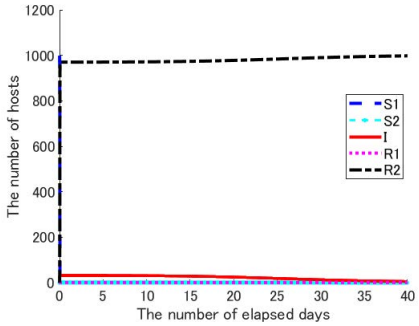

Fig. 7. The number of hosts in each state [$t = 0$ to 40] ($\alpha_1 = \alpha_2 = 0.1$, $\epsilon = \mu = 0.198$).

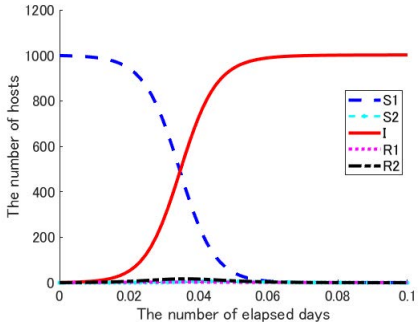

Fig. 8. The number of hosts in each state [$t = 0$ to 0.1] ($\alpha_1 = \alpha_2 = 0.2$, $\epsilon = \mu = 0.1$).

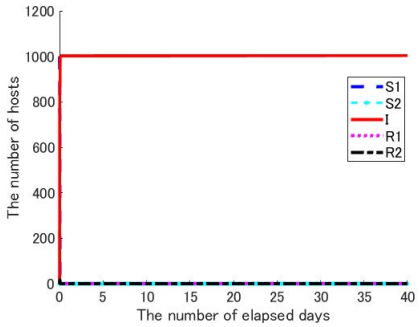

Fig. 9. The number of hosts in each state [$t = 0$ to 40] ($\alpha_1 = \alpha_2 = 0.2$, $\epsilon = \mu = 0.1$).

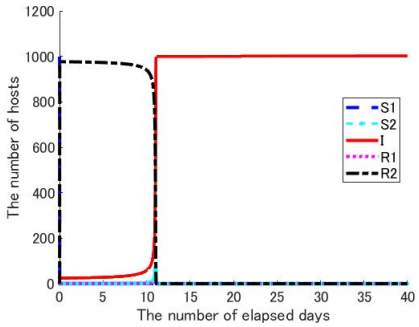

Fig. 10. The number of hosts in each state ($\alpha_1 = \alpha_2 = 0.2$, $\epsilon = \mu = 0.43$).

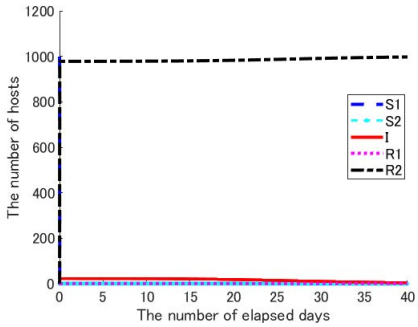

Fig. 11. The number of hosts in each state ($\alpha_1 = \alpha_2 = 0.2$, $\epsilon = \mu = 0.44$).

the number of volunteer hosts. However, joining the volunteer group may degrade their performance because the volunteer hosts should provide a certain amount of their computing resources. Therefore, we should consider this trade-off, using concepts such as the game theory.

ACKNOWLEDGEMENT

This is a product of research which was financially supported by the Kansai University Fund for Supporting Young Scholars, 2018, "Design of anti-malware systems against future malware evolution". This research was partially supported by The Telecommunications Advancement Foundation, Japan.

REFERENCES

[1] J. Dean et al., "Large scale distributed deep networks," in *Proc. Neural Information Processing Systems*, Lake Tahoe, NV, Dec. 2012, pp. 1–11.

[2] G. E. Hinton, S. Osindero, and Y. Teh, "A fast learning algorithm for deep belief nets," *Neural Computation*, vol. 18, no. 7, pp. 1527–1554, 2006.

[3] K. Hongyo, T. Kimura, T. Kudo, Y. Inoue, and K. Hirata, "Modeling of countermeasure against self-evolving botnets," in Proc. *IEEE International Conference on Consumer Electronics - Taiwan (IEEE ICCE-TW 2017)*, Taipei, Taiwan, Jun. 2017, pp. 1–2.

[4] T. Kudo, T. Kimura, Y. Inoue, H. Aman, and K. Hirata, "Stochastic modeling of self-evolving botnets with vulnerability discovery," *Computer Communications*, vol. 124, pp. 101–110, 2018.

[5] Y. Kumai, K. Hongyo, T. Kimura, and K. Hirata, "Infection dynamics of self-evolving botnets with deterministic modeling," in *Proc. the 33rd International Conference on Information Networking (ICOIN 2019)*, Kuala Lumpur, Malaysia, Jan. 2019.

[6] M. A. Rajab, J. Zarfoss, F. Monrose, and A. Terzis, "A multifaceted approach to understanding the botnet phenomenon," in *Proc. ACM SIGCOMM Conference on Internet measurement*, Rio de janeiro, Brazil, Oct. 2006, pp. 1–12.

[7] 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.

[8] 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, pp. 1–10.