World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
Vol:2, No:10, 2008

# Propagation Model for a Mass-Mailing Worm with Mailing List

Akira Kanaoka and Eiji Okamoto



***Abstract*—**Mass-mail type worms have threatened to become a large problem for the Internet. Although many researchers have analyzed such worms, there are few studies that consider worm propagation via mailing lists. In this paper, we present a mass-mailing type worm propagation model including the mailing list effect on the propagation. We study its propagation by simulation with a real e-mail social network model. We show that the impact of the mailing list on the mass-mail worm propagation is significant, even if the mailing list is not large.

***Keywords*—**Malware, simulation, complex networks

## I. Introduction

Mass-mailing type worms that are propagated via e-mail make up a large proportion of all worms. When a user receives e-mail containing such a worm in an attachment file and clicks on it, the computer is compromised. The worm searches for e-mail addresses stored in the computer, and then it sends a massive e-mail-out to all the stored addresses, including that of the worm itself. Although the propagation mechanism is therefore known, the characteristics of this process are still unknown. Recently, research on this topic has expanded, but none of these recent studies has considered the effects of mailing lists on the propagation.

In this paper, we design a mass-mailing type worm propagation mechanism including use of mailing lists, and in a simulation we show that mailing lists exercise a greater effect on the propagating worm than does the network topology.

The rest of this paper is structured as follows. Firstly, we present the background and studies related to our research in Section 2. Then we describe the classical epidemiological models and present a version in matrix form in Section 3. In Section 4, we present an extension of the matrix epidemical model, which can handle mailing lists. In Section 5, we present and discuss the results of a simulation with our epidemical model. In Section 6, we conclude our study and discuss possible future research.

## II. Background and Related Work

### A. Mass-mailing Worms

In this section, we explain how a mass-mailing worm propagates, as a background to our work, and also indicate some related work.

In the case of compromise by a mass-mailing worm, a user first receives an e-mail with an attachment file, which contains the mass-mailing worm. If the user double-clicks on or executes that file, the computer is compromised by the worm.

After compromise, the worm starts to gather e-mail addresses from the address book of the user or from the entire hard disk drive. The basic strategy of the mass-mailing worm is to send e-mails to all addresses gathered from the address book or the hard disk drive. In addition, recent mass-mailing worms have modified their list of addresses by adding well-known account names and removing inconvenient addresses. Such mass-mailing worm extracts all domains from the list of addresses that it has constructed. It then removes all addresses with domains such as government or security company domains. Then the worm adds well-known account names to all domains, and sends itself to every address on its list.

The compromised computer usually begins with a Denial of Service (DoS) attack. Since such an attack launched simultaneously by large numbers of compromised computers results in a Distributed Denial of Service (DDoS) attack, some studies have analyzed trends in network traffic caused by mass-mailing worms (see [8][4]).

Since these studies have considered traffic from computers compromised by mass-mailing worms only, they did not take account of the propagation of such a worm, nor any effect of the mailing-list.

### B. E-mail Network

From the mechanism that we have described, it follows that propagation depends on the topology of the network through which email is sent, which we shall refer to as an *e-mail network*.

Recent research on complex networks has shown that e-mail networks have certain characteristics. Ebel et al. extracted the "From" and "To" e-mail address pairs from SMTP server log files and analyzed that network[3]. If we regard an e-mail address as a node and a From-To pair as a network link, then the degree distribution has the scale-free (power-law) property.

Newman et al. also analyzed e-mail networks[5]. They analyzed address books of the entire user base of their large university computer system, and showed that the degree distribution satisfies an exponential relationship.

We analyzed e-mail networks in a similar fashion, following Ebel's approach. We analyzed our laboratory SMTP server log files over 3 months. There were 31388 mails relayed by the SMTP server, 2550 unique addresses and 3397 unique From-To pairs. The average degree of e-mail addresses was 2.66. Figures 1 and 2 show the in-degree and out-degree distributions. They also show the scale-free property. In this paper, we shall assume that the e-mail network is scale-free.

Akira Kanaoka and Eiji Okamoto are with University of Tsukuba, Japan, email: {kanaoka,okamoto}@risk.tsukuba.ac.jp

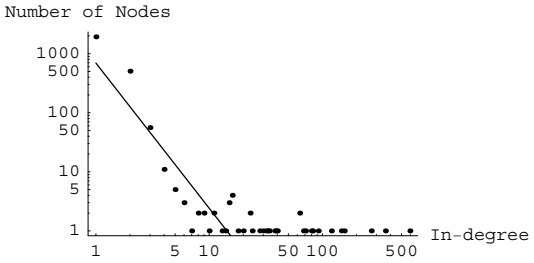


Fig. 1. In-Degree Distribution of addresses on our Laboratory SMTP Server

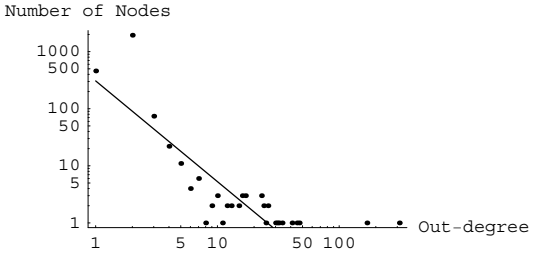


Fig. 2. Out-Degree Distribution of addresses on our Laboratory SMTP Server

Since we propose a mass-mailing worm propagation model which includes a mailing list effect in this paper, we count the number of mailing lists from SMTP server log files. In general, it is difficult to identify mailing list addresses precisely among e-mail addresses. In this study, we extracted 69 mailing list addresses from server logs by using characteristics of the mailing list service tool (Majordomo, fml, Mailman, etc.) settings.

### C. Analysis of Worm Propagation across a Scale-Free Network

Recently, there have been several studies of worm propagation across a scale-free network. Such studies may be divided into two groups, depending on whether the worm studied was a network worm or a mass-mailing worm.

In network worm propagation analysis, researchers made use of the scale-free characteristic of routers or the AS (Autonomous Systems) network topology on the Internet. Briesemeister et al. considered a topology effect for the classical epidemic model and showed how worm propagation was related to it [2]. Nikoloski et al. also considered network worm propagation across a scale-free network [6]. They used a pair approximation technique to characterize the classical epidemic model.

On the other hand, Zou et al. focused on mass-mailing worm propagation across a scale-free network [10] [9]. These studies were focused on the differences between a scale-free network and a random graph, but did not mention any effects of the mailing list on propagation. They used a differential equation for the epidemic model presented by Pastor-Satorras and Vespignani [7]. The equation models infection dynamics of nodes with different degree distributions on the network.

## III. The Classical Epidemic Model in terms of Matrices

### A. Requirements for Model Extension

The classical epidemic model cannot be applied directly to yield a model, which includes a mailing list effect. A worm propagation model that includes a mailing list effect has to satisfy the following three conditions:

*a) 1) Expression of Flexible Network Topology:* Although e-mail networks have been assumed to be scale-free in research analyzing the propagation of mass-mailing worms, some studies have shown that an e-mail network may involve an exponential distribution. To analyze how a mass-mailing worm propagates through a real network, it is necessary for the model to be able to handle a flexible network topology.

*b) 2) State Transition Induced by Anti-virus Software:* In classical epidemic models, a state transits from susceptible (S) to infected (I), then transits from infected (I) to removed (R). Since anti-virus software guards against worm/virus infection, whether the computer is infected or not, it is necessary for the model to deal with state transitions from S to R.

*c) 3) Mailing List Effect for Worm Propagation:* E-mail addresses may be regarded as nodes in a network; however, mailing list e-mail address nodes behave differently from normal e-mail address nodes. A mailing list is a kind of amplifier of an e-mail message, whether or not the message contains a worm. Intuitively it seems that such amplification makes propagation more rapid. Therefore it is necessary for the model to deal with mailing lists.

In the following subsections, we describe the classical SI/SIR model

### B. SI/SIR Model

The classical epidemic model was created by Kermack and McKendrick. There are three states: Susceptible (S), Infected (I), and Removed (R). The number of nodes in each state is described by differential equations as follows:

$$\begin{aligned} \frac{dS}{dt} &= -\beta SI \\ \frac{dI}{dt} &= \beta SI - \gamma I \\ \frac{dR}{dt} &= \gamma I \end{aligned} \tag{1}$$

In these equations, the infection rate is denoted by $\beta$ and the removal rate is denoted by $\gamma$. Without state R, the model (SI model) is represented by $\frac{dS}{dt} = -\beta SI, \frac{dI}{dt} = \beta SI$.

Since they use the term "differential" equation even though its values ($S$, $I$, $R$) are discrete in these equations, we also use the word "differential" in this paper.

These equations are based on a complete network, but the network topology is not considered.

### C. SIS Model

There is another model for epidemics, called the SIS model. In the SIS model, a node in state "I" returns to "S" instead of "R" in the SIR model.

$$\frac{dS}{dt} = -\beta SI + \gamma I$$
$$\frac{dI}{dt} = \beta SI - \gamma I \tag{2}$$

Pastor-Satorras and Vespignani expand this model by allowing the infection dynamics to be related to the degrees of the nodes [7]. They include an equation of the probability $\rho_k(t)$ that a node with $k$ links is infected.

$$\frac{d\rho_k(t)}{dt} = -\rho_k(t) + \beta k\,(1-\rho_k(t))\,\Theta(\beta) \tag{3}$$

$\Theta(\beta)$ is the probability that any given link points to an infected node and is described as follows, using $P(k)$, which is the fraction of nodes that have degree $k$ :

$$\Theta(\beta) = \sum_k \frac{kP(k)\rho_k}{\sum_s sP(s)} \tag{4}$$

Although this differential equation is employed by Zou et al ([10],[9]), this model describes the dynamics of all kinds of worm because of its use of the SIS model. We cannot employ this model directly, since we want to focus on specific worm epidemic dynamics. Furthermore, this model does not consider the effects of mailing lists on worm multiplication.

### D. Expression of the SI/SIR Model in terms of Matrices

The classical SI/SIR models and other studies as discussed above do not satisfy the conditions of Section 3.1. To derive a model satisfying the conditions, we present an alternative description of the epidemic model, which makes use of adjacency matrices.

Some definitions of notation and state transition expressions follow:

*d) Some notation:*

| | | |
|---|---|---|
| $\mathbf{A} = \{a_{xy}\}$ | : | $M \times N$ matrix $\mathbf{A}$ and its element |
| $\mathbf{E}$ | : | Identity matrix |
| $\mathbf{AB}$ | : | Matrix multiplication |
| $\mathbf{A} \cdot \mathbf{B}$ | : | Each element multiplication |

$$F(\mathbf{A}) = \{f(a_{xy})\}$$

$$f(a_{xy}) = \begin{cases} 1 & \text{if } a_{xy} \geq 0.5 \\ 0 & \text{else} \end{cases}$$

*e) State Matrix:*

$$\mathbf{I}(t) = \{i_x(t)\}$$

$$i_x(t) = \begin{cases} 1 & \text{if node } x \text{ is infected at time } t \\ 0 & \text{else} \end{cases}$$

$$\mathbf{R}(t) = \{r_x(t)\}$$

$$r_x(t) = \begin{cases} 1 & \text{if node } x \text{ is removed at time } t \\ 0 & \text{else} \end{cases}$$

If $i_x(t) = 0$ and $r_x(t) = 0$, the node $x$ is in state "S".

*Expression of the Network:*

$$\mathbf{T} = \{t_{xy}\}$$

$$t_{xy} = \begin{cases} 1 & \text{if } x \to y \text{ has a link} \\ 0 & \text{else} \end{cases}$$

If $\mathbf{T}$ is a symmetric matrix, i.e. $t_{xy} = t_{yx}$, the network defined by $\mathbf{T}$ is an undirected graph.

*State Transition of* $\mathbf{R}(t)$*:*

$$\mathbf{R}(t+1) = F\left(\mathbf{R}(t) + \mathbf{I}(t) \cdot \mathbf{\Lambda}\right) \tag{5}$$
$$r_x(t+1) = f\left(r_x(t) + i_x(t)\lambda_x(t)\right) \tag{6}$$

$$\mathbf{\Lambda} = \{\lambda_x(t)\}$$

$$\lambda_x(t) = \begin{cases} 1 & \text{if the random value } \tau \leq \gamma \\ 0 & \text{otherwise} \end{cases}$$

*State Transition of* $\mathbf{I}(t)$*:*

$$\mathbf{I}(t+1) = F\left(\mathbf{I}(t)\left(\mathbf{E} + \mathbf{D}(t) \cdot \mathbf{T}\right)\right) \cdot \left(\mathbf{1} - \mathbf{R}(t)\right) \tag{7}$$

$$i_x(t+1) = f\left(\sum_y i_y(t)(e_{yx} + d_{yx}(t)t_{yx})\right)(1 - \lambda_x(t)) \tag{8}$$

$$\mathbf{D} = \{d_{xy}(t)\}$$

$$d_{xy}(t) = \begin{cases} 1 & \text{if the random value } \tau \leq \beta \\ 0 & \text{otherwise} \end{cases}$$

### E. Matrix Expression of the Classical Model

In this section, we show our matrix expression can be chosen to match the classical model, thus showing the correctness of our model.

Let $S(t)$ (respectively $I(t)$, $R(t)$) denote the number of nodes that are in state $S$ (respectively $I$, $R$). Then transitions between states may be expressed as follows:

$$\begin{aligned} S(t+1) &= S(t) - S(t)p(S \to I) \\ I(t+1) &= I(t) + S(t)p(S \to I) - I(t)p(I \to R) \\ R(t+1) &= R(t) + I(t)p(I \to R) \end{aligned} \tag{9}$$

Now, $S \to I$ means ($i_x(t) = 0$ and $\lambda_x(t) = 0$) $\to$ ($i_x(t+1) = 1$ and $\lambda_x(t+1) = 0$). Then it goes to (any $i_k(t)d_{kx}(t)t_{kx} = 1$ at $1 \leq k \leq N$), and $p(S \to I)$ as follows:

$$\begin{aligned} p(S \to I) &= p(\text{any } i_y(t)d_{yx}(t)t_{yx} = 1) \\ &= 1 - (1-\beta)^{\sum_y i_y(t)t_{yx}} \\ &= 1 - (1-\beta)^{m_x(t)} \end{aligned} \tag{10}$$

$m_x(t) = \sum_y i_y(t)t_{yx}$ means the number of infected nodes which have a link to node $x$. In the case of a complete network, since all $t_{kx} = 1$,

$$m_x(t) = \sum_y i_y(t) = I(t)$$

and, if $\beta \ll 1$, Eq. (10) goes to

$$p(S \to I) = \beta I(t) \tag{11}$$

Also, $p(I \to R)$ is as follows:

$$p(I \to R) = \gamma \tag{12}$$

From Eq. (11) and (12), Eq. (9) can be changed as follows.

$$\begin{aligned} S(t+1) &= S(t) - \beta S(t) I(t) \\ I(t+1) &= I(t) + \beta S(t) I(t) - \gamma I(t) \\ R(t+1) &= R(t) + \gamma I(t) \end{aligned} \tag{13}$$

When we also change the Eq. (13) expression to differential equation form, it can be seen that it is the same as Eq. (1).

## IV. Epidemic Model for Mass-mailing Worm

### A. Variant SIR Model for Mass-mailing Worm

Our model above satisfies the first requirement. Thus, we need to modify the model to satisfy the second requirement. This is quite easy. We merely need to adjust equation (5) as follows:

$$\mathbf{R}(t+1) = F\left(\mathbf{R}(t) + \mathbf{\Lambda}\right) \tag{14}$$

### B. Mailing List Expression

In this section, we extend the matrix expression of the epidemic described in III-D to satisfy the third requirement.

To handle mailing lists (ML), we treat an ML address as a special node. When an ML node receives an e-mail from another node, it forwards the e-mail to all ML members immediately. On the other hand, when a normal user node receives an e-mail (in this case, a mail containing a mass-mailing worm), the mail is not forwarded to other nodes until the user is infected, i.e. the user activates the attachment file in the mail. An ML node has a 2-step action (infection and propagation) whereas a normal node has a 1-step action (Fig. IV-B).

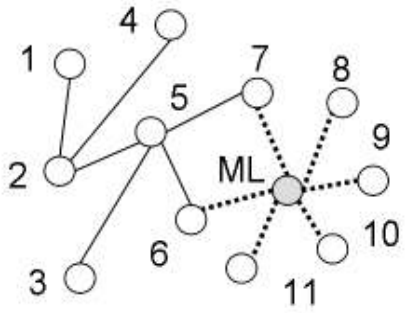


Fig. 3. Treating a Mailing List Address as a Special Node

To allow for the special characteristics of an ML node, we firstly convert the expression of the time scale. When the time value is even, the action of a normal node and the first action of an ML node (infection) are carried out. The second action of an ML node (propagation) is carried out in odd time steps.

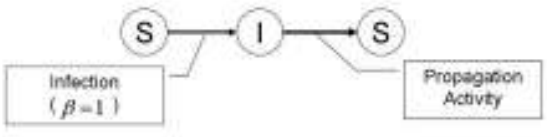


Fig. 4. 2 Step Epidemic Dynamics of an ML Node

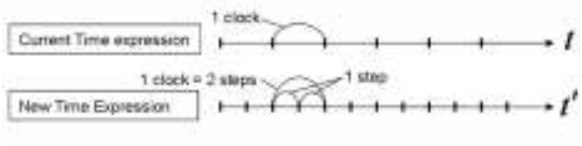


Fig. 5. Time Scale for the Proposed Model

Next we extend the matrix expression to deal with an ML node. The state matrix $\mathbf{I}(t)$, $\mathbf{R}(t)$ and the network expression $\mathbf{T}$ are extended to $\mathbf{I}'(t)$, $\mathbf{R}'(t)$, $\mathbf{T}'$, using the ML node state matrix $\mathbf{I}_{\{ML\}}(t)$, $\mathbf{R}_{\{ML\}}(t)$ and the network topology $\mathbf{T}_{\{U\to ML\}}, \mathbf{T}_{\{ML\to U\}}, \mathbf{T}_{\{ML\to ML\}}$ which denote the topologies of user nodes to ML nodes, ML nodes to user nodes and ML nodes to ML nodes respectively. Thus, $\mathbf{0}$ is a matrix with all elements equal to 0, and $\mathbf{1}$ is a matrix with all elements equal to 1.

$$\begin{aligned} \mathbf{I}'(t) &= \left\{\mathbf{I}(t), \mathbf{I}_{\{ML\}}(t)\right\} \\ \mathbf{R}'(t) &= \left\{\mathbf{R}(t), \mathbf{R}_{\{ML\}}(t)\right\} \\ \mathbf{T}' &= \begin{pmatrix} \mathbf{T} & \mathbf{T}_{\{U\to ML\}} \\ \mathbf{T}_{\{ML\to U\}} & \mathbf{T}_{\{ML\to ML\}} \end{pmatrix} \\ \hat{\mathbf{T}}_{\{U\to\}} &= \left\{\mathbf{T}, \mathbf{T}_{\{U\to ML\}}\right\} \\ \hat{\mathbf{T}}_{\{ML\to\}} &= \left\{\mathbf{T}_{\{ML\to U\}}, \mathbf{T}_{\{ML\to ML\}}\right\} \end{aligned}$$

Then, we also extend state transitions to $\mathbf{R}'(t)$ and $\mathbf{I}'(t)$. $\mathbf{I}'(2t+1)$ is extended from Eq(7).

$$\begin{aligned} \mathbf{R}'(2t+1) &= F\left(\mathbf{R}'(2t) + \{\mathbf{\Lambda}, \mathbf{0}\}\right) \\ \mathbf{R}'(2t+2) &= \mathbf{R}'(2t+1) \end{aligned} \tag{15}$$

$$\begin{aligned} &\mathbf{I}'(2t+1) = \\ &\quad F\left(\mathbf{I}'(2t)\left(\{\mathbf{E}, \mathbf{0}\} + \mathbf{D}'(2t)\cdot\hat{\mathbf{T}}_{\{U\to\}}\right)\right) \\ &\quad \cdot\left(\{\mathbf{1} - \mathbf{R}(2t), \mathbf{0}\}\right) \end{aligned} \tag{16}$$

$$\begin{aligned} &\mathbf{I}'(2t+2) = \\ &\quad F\left(\{\mathbf{I}'(2t+1), \mathbf{0}\} + \mathbf{I}_{\{ML\}}(2t+1)\hat{\mathbf{T}}_{\{ML\to\}}\right) \\ &\quad \cdot(\mathbf{1} - \mathbf{R}'(2t+2)) \end{aligned} \tag{17}$$

### C. Differential Equations for the Proposed Model

In this section, we provide differential equations for the variant SIR model and the mailing list model. These equations are derived in the same manner as those in section III-E.

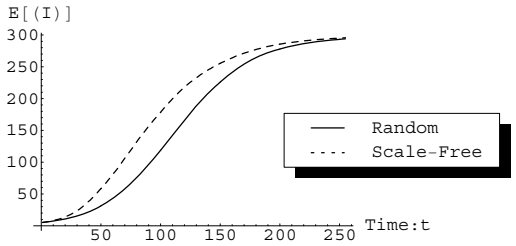


Fig. 6. SI Model Simulation : $\beta = 0.005, E[k] = 8$

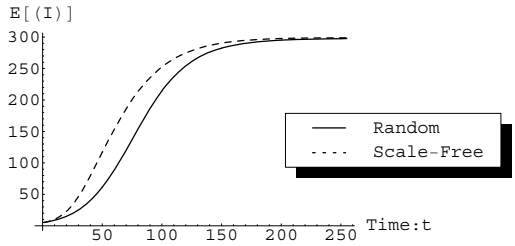


Fig. 7. SI Model Simulation : $\beta = 0.01, E[k] = 8$

*1) Variant SIR Model for Mass-mailing Worm:*

$$\frac{dS}{dt} = -(1-\gamma)\beta S m_x(t) - \gamma S$$
$$\frac{dI}{dt} = (1-\gamma)\beta S m_x(t) - \gamma I \quad (18)$$
$$\frac{dR}{dt} = \gamma(S+I)$$

*2) Mailing List Model:*

$$\frac{dS}{dt} = -(1-\gamma)\beta S\Omega - \gamma S$$
$$\frac{dI}{dt} = (1-\gamma)\beta S\Omega - \gamma I \quad (19)$$
$$\frac{dR}{dt} = \gamma(S+I)$$

where $\Omega = m_x(2t) + \sum_{y=N+1}^{N+M} m_y(2t)t_{yx}$.

## V. Simulation

We simulate each model described above and compare the results with those of previous studies. We focus on three aspects for comparison: 1) propagation differences due to the network topology, 2) comparisons between the SIR model and the variant SIR model, and 3) the mailing list effect on worm propagation.

The number of nodes $N$ was taken as 300 and the number of initially infected nodes $I(0)$ was taken as 5 in all simulations. We use a scale-free topology generation algorithm from work of Barabasi and Albert [1]. Thus, we take $\mathbf{T}_{\{ML \to ML\}} = \mathbf{0}$, that is, we assume that members of one mailing list are not included in other mailing lists in our simulations.

### A. Scale-free Topology Effects on Propagation

At first, we study propagation differences between scale-free networks and random graphs.

For a start, we compare transitions of $I(t)$ for the SI model to measure simple speed of propagation. In the simulation, we generated both networks so that both were of the same average degree $E[k]$. Fig. 6 shows the result of the simulation for the infection rate $\beta = 0.005$ and average degree $E[k] = 8$, and Fig. 7 shows the result for $\beta = 0.01$ and $E[k] = 8$. Both results indicate that a scale-free network leads to more rapid worm propagation than a random graph.

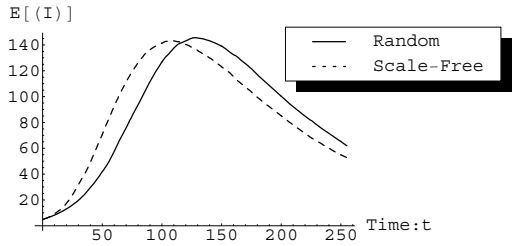


Fig. 8. SIR Model Simulation : $\beta = 0.005, \gamma = 0.01, E[k] = 12$

Next, we compare transitions of $I(t)$ for the SIR model. Fig. 8 shows the result of a simulation with infection rate $\beta = 0.005$, removal rate $\gamma = 0.01$, and average degree $E[k] = 12$, and Fig. 9 shows the result with $\beta = 0.01$, $\gamma = 0.01$ and $E[k] = 12$. For this model also, worms propagated more rapidly across a scale-free network than across a random graph. Furthermore, peaks of $I(t)$ in a scale-free network occurred earlier than for a random graph.

### B. Evaluation of the Variant SIR Model

Secondly, we studied $I(t)$ transitions for the variant SIR model and compared it with the SIR model. We used a scale-free network with $E[k] = 12$. Fig. 10 shows the result with $\beta = 0.005$, $\gamma = 0.01$, and Fig. 11 shows the result with $\beta = 0.01$ and $\gamma = 0.01$.

The variant SIR model resulted in a low peak for both cases. Basically, the cause of the low peaks in the variant SIR model is clearly the effect of the $S \to R$ state transitions. The peak

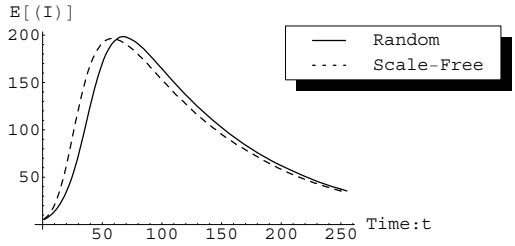


Fig. 9. SIR Model Simulation : $\beta = 0.01, \gamma = 0.01, E[k] = 12$

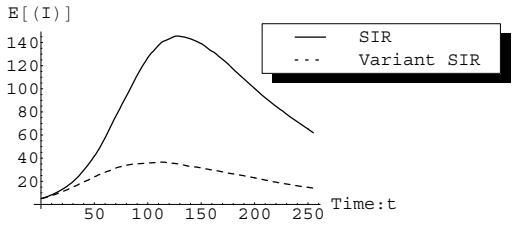


Fig. 10. SIR Model and Variant SIR Model: $\beta = 0.005, \gamma = 0.01$

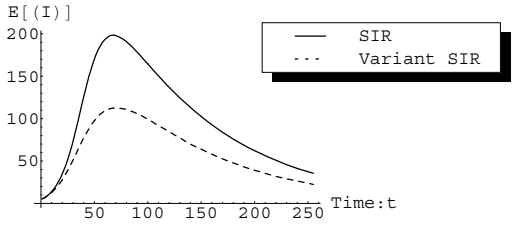


Fig. 11. SIR Model and Variant SIR Model: $\beta = 0.01, \gamma = 0.01$

values in Fig 10 were 143.23 at $t = 107$ for the SIR mode and 54.86 at $t = 90$ for the variant SIR mode. The ratio of the peak values was 0.38. Similarly the peak values in Fig 11 were 196.54 at $t = 59$ for the SIR mode and 130.53 at $t = 60$ for the variant SIR mode. The ratio of the peak values was 0.66.

Because the removal rate $\gamma$ was the same for both, the difference in peak values was due only to the infection rate $\beta$.

### C. Mailing List Impact on Mass-mailing Worm Propagation

In this section, we compare the ML model and the variant SIR model with respect to the mailing list effect on worm propagation. We prepared two types of network for use with the ML model. The first network contained 3% ML nodes; the second network contained 5% ML nodes. $E[k] = 12$ for both networks and both had a scale-free network topology.

Fig. 12 shows the result with $\beta = 0.005$, $\gamma = 0.01$. We call this Case 1. Fig. 13 shows the result with $\beta = 0.01$, $\gamma = 0.01$. We call this Case 2.

We can see quite rapid propagation and high peak for $I(t)$ in both cases. Although we found differences for the peak ratios in section V-B and the cause can be considered to be due to differences in the infection rate, these peaks are quite high. It appears that the peak value is affected more strongly by the proportion of ML nodes than the infection or removal rate.

Furthermore, the most notable feature of propagation using the ML model is its growth in the first three steps. We can see the outbreak with the ML model in Table V-C, which shows each value at $t' = 0, 1, 2$ and 3. Figures 14, and 15 also show this. These differences are caused by the increasing value of the factor $\Omega$ in $\frac{dI}{dt}$ of Eq.19. It seems that a hub node of large degree $k$ infected at an earlier time by an email posted to a mailing list caused the outbreak.

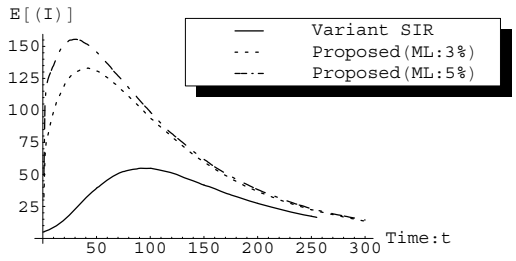


Fig. 12. Mailing List Impact: $\beta = 0.005$, $\gamma = 0.01$

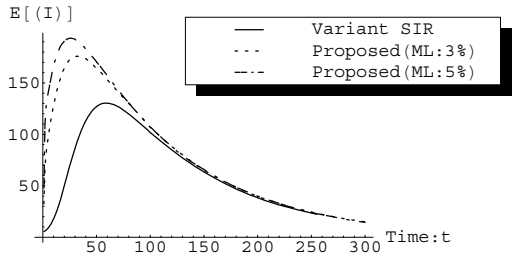


Fig. 13. Mailing List Impact: $\beta = 0.01$, $\gamma = 0.01$

## VI. Conclusion

In this paper, we have proposed a new propagation model for the mass-mailing worm. Our model includes a mailing list effect on propagation and such an effect has not previously been considered. We simulate mass-mailing worm propagation using the proposed model and show that the mailing list has a large effect on the propagation. Although previous studies have mainly considered propagation characteristics with respect to a scale-free network topology, our results show that mailing lists have a more powerful effect on worm propagation.

There remains more work to do on modeling mass-mailing worm propagation. For example, a representative mass-mailing worm NetSky still exists in the wild, though its outbreak occurred in 2004. Such survival can be compared to survival of a real virus in the world. To express survival of a mass-mailing worm, we have to consider adding more conditions to this model, such as periodic propagation activities of an infected node, and re-infection of a removed node. Also, determining the appropriate differential equations for a proposed model such as that expressed by Eq.(3) will enrich the study of this topic.

TABLE I
$I(t)$ Transition

| $t'$ | 0 | 1 | 2 | 3 |
|---|---|---|---|---|
| Case 1(variant SIR) | 5 | 5.3 | 5.57 | 5.83 |
| Case 1(ML:3%) | 5 | 32.43 | 63.83 | 74.17 |
| Case 1(ML:5%) | 5 | 49.12 | 104.61 | 117.36 |
| Case 2(variant SIR) | 5 | 5.3 | 5.57 | 5.83 |
| Case 2(ML:3%) | 5 | 32.48 | 67.92 | 81.44 |
| Case 2(ML:5%) | 5 | 47.74 | 109.88 | 126.6 |

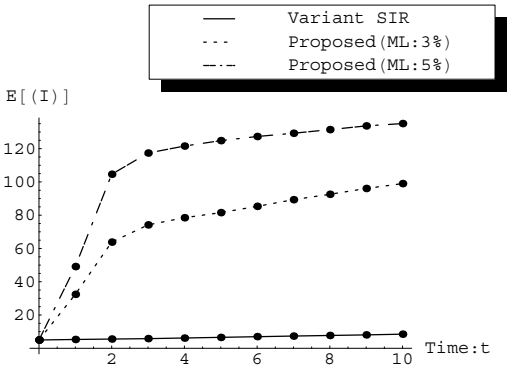


Fig. 14. First Ten Steps in Fig. 12

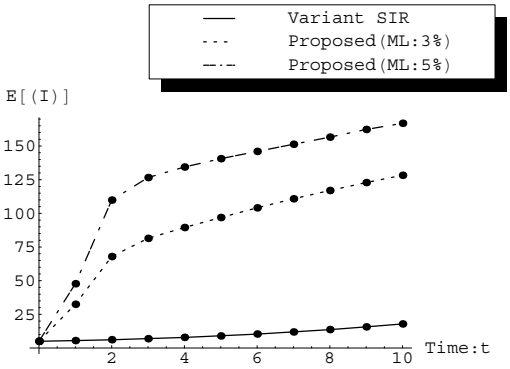


Fig. 15. First Ten Steps in Fig. 13

## REFERENCES

[1] A. Barabási and R. Albert. Emergence of scaling in random networks. *Science*, 286:509–512, 1999.
[2] L. Briesemeister, P. Lincoln, and P. Porras. Epidemic profiles and defense of scale-free networks. In *Proc. of ACM Workshop on Rapid Malcode (WORM'03)*, 2003.
[3] H. Ebel, L.-I. Mielsch, and S. Bornholdt. Scale-free topology of e-mail networks. In *PHYSICAL REVIEW E 66, 035103(R)*, 2002.
[4] K. Ishibashi, T. toyono, and K. Toyama. Detecting mass-mailing worm infected hosts by mining dns traffic data. In *Proc of ACM SIGCOMM 2005 Workshop on Mining Network Data (MineNet2005)*, 2005.
[5] M. E. J. Newman, S. Forrest, and J. Balthrop. Email networks and the spread of computer viruses. In *PHYSICAL REVIEW E 66, 035101(R)*, 2002.
[6] Z. Nikoloski, N. Deob, and L. Kucera. Correlation model of worm propagation on scale-free networks. *COMPLEXUS*, VOL. 3(1-3):pp.169–182, 2006.
[7] R. Pastor-Satorras and A. Vespignani. Epidemic spreading in scale-free networks. *Physical Rev. Letters*, vol. 86, 2001.
[8] C. Wong, S. Bielski, J. M. McCune, and C. Wang. A study of mass-mailing worms. In *Proc. of ACM Workshop on Rapid Malcode (WORM'03)*, 2003.
[9] C. C. Zou, D. F. Towsley, and W. Gong. Email worms modeling and defense. In *ICCCN*, pages 409–414, 2004.
[10] C. C. Zou, D. F. Towsley, and W. Gong. Modeling and simulation study of the propagation and defense of internet e-mail worms. *IEEE Trans. Dependable Sec. Comput.*, 4(2):105–118, 2007.

**Akira Kanaoka** He received B.S. and M.S. in information science from Toho University in 1998 and 2001 respectively. He received Ph.D in computer science from University of Tsukuba in 2004. He worked for SECOM Co., Ltd. since 2004. From 2007 he worked as a postdoctoral fellow at University of Tsukuba. He is working as an assistant professor at University of Tsukuba. His research interests are network security and electronic authentication.

**Eiji Okamoto** Professor Eiji Okamoto received his B.S., M.S. and Ph.D degrees in electronics engineering from the Tokyo Institute of Technology in 1973, 1975 and 1978, respectively. He worked and studied communication theory and cryptography for NEC central research laboratories since 1978. From 1991 he became a professor at Japan Advanced Institute of Science and Technology, then at Toho University. Now he is a professor at Graduate School of Systems and Information Engineering, University of Tsukuba. His research interests are cryptography and information security. He is a member of IEEE and a coeditor-in-chief of International Journal of Information Security.