



ELSEVIER

Available online at www.sciencedirect.com



August 2010, 17(4): 63–68
www.sciencedirect.com/science/journal/10058885

The Journal of China
Universities of Posts and
Telecommunications

<http://www.jcupt.com>

Modeling and analysis of Internet worm propagation

SU Fei¹(✉), LIN Zhao-wen¹, MA Yan^{1,2}

1. Institute of Networking Technology, Beijing University of Posts and Telecommunications, Beijing 100876, China

2. Beijing Key Laboratory of Intelligent Telecommunications Software and Multimedia, Beijing University of Posts and Telecommunications, Beijing 100876, China

Abstract

Although the frequency of Internet worm's outbreak is decreased during the past ten years, the impact of worm on people's privacy security and enterprise's efficiency is still a severe problem, especially the emergence of botnet. It is urgent to do more research about worm's propagation model and security defense. The well-known worm models, such as simple epidemic model (SEM) and two-factor model (TFM), take all the computers on the internet as the same, which is not accurate because of the existence of network address translation (NAT). In this paper, we first analyze the worm's functional structure, and then we propose a three layer worm model named three layers worm model (TLWM), which is an extension of SEM and TFM under NAT environment. We model the TLWM by using deterministic method as it is used in the TFM. The simulation results show that the number of NAT used on the Internet has effects on worm propagation, and the more the NAT used, the slower the worm spreads. So, the extensive use of NAT on the Internet can restrain the worm spread to some extent.

Keywords worm propagation, model, TLWM, NAT

1 Introduction

A computer worm is a self-replicating computer program. It uses network to send copies of itself to other nodes (computers on the network) and it may do so without any user intervention. Unlike a virus, it does not need to attach itself to an existing program.

The first known worm appeared in the wild was the Morris worm in 1988. Since then, new worms appeared frequently [1]. In 2001, the Code Red and Nimda worms quickly infected hundreds of thousands of computers, causing millions of dollars loss to our society [2]. The slammer worm appeared on January 25th, 2003, and quickly spread throughout the Internet. Thereby, the security threats caused by network worms have increased dramatically. In 2007, The Storm Worm began infecting thousands of (mostly private) computers in Europe and the United States. In 2008, the worm Conficker was detected. It is now believed to be the largest computer worm infection since the 2003 SQL Slammer.

A functional division of Internet worm is mainbody

function and auxiliary function [2]. The mainbody function module is composed of four parts. That is information collection module, probe module, attack module and self-propagating module. The information collection module decides which algorithm should be adopted by the worm to search information on local and remote network. The content of the information includes local system information, user information, mailing list, border router's information and so on. The information can be shared by other individuals, solely. Probe module is responsible for the host's detection of fragility, and then decides which attack mode should be taken. The attack module uses obtained security vulnerability to build propagation route. Such module is open and extensible on attack method. The self-propagating module can use different measures to generate different worm copy, and transmit the worm copy to different hosts.

It is necessary to study the worm's spread process from different viewpoints. So, an analytical worm propagation model is needed. The purpose of the worm model is to identify the weakness in the worm spreading chain and provides accurate defending measure for the epidemiology research. We can study the worm's behavior by an accurate

Received date: 21-05-2009

Corresponding author: SU Fei, E-mail: suf@buptnet.edu.cn

DOI: 10.1016/S1005-8885(09)60489-1

and effective worm model. The large-scale worm infestations appeared in the last several years have triggered several efforts to model worm spread. A majority of worm propagation models are based on deterministic epidemic models [3]. It can simulate the worm propagation in large-scale network environment well. It derives from the epidemiology area. In epidemiology area, both stochastic models and deterministic models exist for modeling the spreading of infectious diseases. Stochastic models are suitable for small-scale network with simple virus dynamics, while deterministic models are suitable for large-scale network under the assumption of mass action, relying on the law of large number. Here, we only consider deterministic model for worm propagation. And we do not take the network topology into account, although it is an important aspect to research the worm propagation.

2 Related work

In the past years, some excellent and novel worm models have been proposed. Zou et al. analyze two typical worm models SEM and Kermack-Mckendrick (KM) model, and propose a new model named two-factor worm model [3] which takes human countermeasure and network congestion into account. Later, Zou et al. analyze different scan strategies affecting worm's propagation [4]. Chen et al. present a new deterministic approximation model, named analytical active worm propagation (AAWP) model by probability theory [5]. Su et al. propose a worm model by dividing the Internet into different groups based on TFM [6], and then they analyze the effect of firewall on the propagation of Internet worm [7–8]. Wang et al [9]. study the worm propagation from the eigenvalue viewpoint, and propose a threshold value of worm outbreak. Okamura present a Markovian model method to analyze worm's spread [10]. Kamra A et al. discuss worm propagation in IPv6 work, and study the effect of DNS delays on worm propagation [11].

Most of the model considered the worm's spread from the overall point of view. Few models calculate the local factors' effects to the worm propagation, such as firewall, NAT and intrusion detection system (IDS). With the extensive use of these devices, it more or less affects the worm's propagation. Xing et al. propose a worm model under NAT environment based on AAWP worm model [12], which can only reflect the worm propagation trend. We cannot get the effects of NAT on the worm spread from this model.

3 Epidemic model introduction

In the epidemiology research, the mathematic theory concerning the infectious disease is well understood. As to the similarity of infectious disease and worm propagation, we can use the theory to study the spread of Internet worm. In this section, we briefly introduce three kinds of epidemic models.

According to epidemiology modeling, hosts that can be infected by worm are called susceptible hosts; hosts that have been infected and can infect others are called infectious hosts; hosts that are immune such that they cannot be infected are called removed hosts. In this paper, we will use the same terminology for Internet worm modeling.

We list the notations used in this paper as follows:

1) $S(t)$: the number of susceptible hosts at time t .

2) $I(t)$: the number of infectious hosts at time t .

3) $R(t)$: the number of removed hosts from infectious hosts at time t .

4) $Q(t)$: the number of removed hosts from susceptible hosts at time t .

5) N : the number of hosts in the system.

6) $J(t)$: the number of infectious hosts including removed hosts at time t .

7) $\beta(t)$: the infection rate.

3.1 Simple epidemic model

In simple epidemic model, each host stays in one of two states: susceptible or infectious. The model assumes that the system is homogeneous—each host has the equal probability to contact any other hosts in the Internet. Once a host is infected by a worm, it remains in the infectious state forever. Thus the number of contacts between infectious hosts and susceptible hosts is proportional to $S(t)I(t)$. Based on assumptions, the simple epidemic model for a finite population is

$$\frac{dI(t)}{dt} = \beta I(t)[N - I(t)] \quad (1)$$

where β is called the pairwise rate of infection. At $t=0$, $I(0)$ hosts are infectious and the other $S(0) = N - I(0)$ hosts are all susceptible.

3.2 KM model

KM model takes the removal process of infectious hosts into consideration. It assumes that during an epidemic of a contagious disease, some individuals can be either recover or die, and thus they are immune to the disease forever.

Therefore, in this model each host stays in one of three states at any time: susceptible, infectious, or removed. Each host either makes the state transition ‘susceptible → infectious → removed’ or remains in ‘susceptible’ state all the time.

Based on the simple epidemic model Eq. (1), the KM model is

$$\left. \begin{aligned} \frac{dI(t)}{dt} &= \beta I(t)[N - J(t)] \\ \frac{dR(t)}{dt} &= \gamma I(t) \\ J(t) &= I(t) + R(t) = N - S(t) \end{aligned} \right\} \quad (2)$$

where γ is the removal rate of the infectious hosts.

3.3 Two-factor model

Former worm models neglect the dynamic effect, such as human countermeasures on worm behavior and the change of infection rate during infection. The two-factor worm model considers the two factors: human countermeasures and decreased infection rate.

In order to consider the two factors, the change in the number of susceptible hosts $S(t)$ from time t to time $t + \Delta t$ follows the equation:

$$\frac{dS(t)}{dt} = -\beta(t)S(t)I(t) - \frac{dQ(t)}{dt} \quad (3)$$

and the infectious hosts from t to $t + \Delta t$ are:

$$\frac{dI(t)}{dt} = \beta(t)[N - R(t) - I(t) - Q(t)]I(t) - \frac{dR(t)}{dt} \quad (4)$$

From the analysis above, we know that the two-factor worm model is more accurate. When the infection rate $\beta(t)$ becomes constant and do not consider the removal process from susceptible population, i.e., $Q(t)=0$, the two-factor worm model can be degenerated to KM worm model. $\beta(t)$, $Q(t)$ and $R(t)$ are dynamic factors. The equations of $\beta(t)$, $R(t)$, $Q(t)$, $S(t)$ are as follows:

$$\left. \begin{aligned} \beta(t) &= \beta_0 \left[1 - \frac{I(t)}{N} \right]^\eta \\ \frac{dR(t)}{dt} &= \gamma I(t) \\ \frac{dS(t)}{dt} &= -\beta(t)S(t)I(t) - \frac{dQ(t)}{dt} \\ \frac{dQ(t)}{dt} &= \mu S(t)J(t) \end{aligned} \right\} \quad (5)$$

Using the equations above, the complete two-factor worm model is:

$$\left. \begin{aligned} \frac{dS(t)}{dt} &= -\beta(t)S(t)I(t) - \frac{dQ(t)}{dt} \\ \frac{dR(t)}{dt} &= \gamma I(t) \\ \frac{dQ(t)}{dt} &= \mu S(t)J(t) \\ \beta(t) &= \beta_0 \left[1 - \frac{I(t)}{N} \right]^\eta \\ N &= S(t) + I(t) + R(t) + Q(t) \\ I(0) &= I_0 \ll N; S(0) = N - I_0; R(0) = Q(0) = 0 \end{aligned} \right\} \quad (6)$$

The dynamic curves of $I(t), J(t), S(t)$ are shown in Fig. 1.

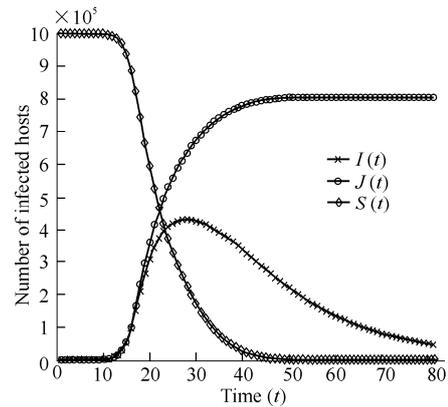


Fig. 1 Two-factor model

4 Effect of NAT on worm propagation

4.1 Influence of NAT on worm propagation

Network address translation is a mechanism of replacing IP address information in packet headers while in transit across a traffic routing device for the purpose of remapping a given address space into another. Nowadays, it is widely used in IPv4 network. However, most of the worm models do not take NAT into account. They assume that all the hosts on the Internet are the same. Infected hosts in the system can reach any vulnerable host directly. However, this is not the case at all. Considering the hosts behind the NAT, an infected host cannot infect this kind of hosts directly, because it cannot get its accurate IP address. The IP address of the hosts behind NAT is private. We call this kind of host ‘inner host’. The hosts except inner hosts and NAT hosts are outer hosts. To compromise the inner hosts, the infected host must infect the NAT host first, and then, the NAT hosts can use some scan strategies to infect the inner hosts. Therefore, if we model the behavior of worm propagation, NAT hosts should be taken into consideration to improve the accuracy of the model.

The structure of the propagation model in this case is

shown in Fig. 2. The rectangle represents the NAT hosts. The black circle is the inner hosts. Other circles denote the routers and outer hosts.

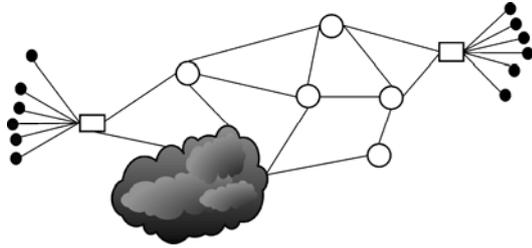


Fig. 2 The structure of NAT environment

4.2 Worm model under NAT environment

From Fig. 2 we can easily understand the worm propagation with the existence of NAT. In this section, we use differential equation to model the worm propagation under this condition based on simple worm model and two-factor worm model. Here, we make some assumptions about this model. The hosts and routers except the inner hosts are outer hosts. The NAT hosts are also taken as outer hosts. We assume that the NAT hosts distribute in the propagation system homogeneously. The NAT hosts are infected at discrete time during the worm propagation. We assume that the time at which the NAT hosts be infected is uniformly distributed after the worm's outbreak. The worm propagation in inner hosts follows simple worm model.

We model the worm propagation under NAT environment named three layers worm model (TLWM). The model divides the Internet hosts into three layers. The outer hosts belong to the first layer. The NAT hosts and inner hosts compose the second and the third layer respectively. In the first layer, the worm spread on the Internet without considering NAT hosts and inner hosts. We can model this case by simple worm model or two-factor worm model. The notations in this layer are the same as the notations in the two models. The second layer describes the worm propagation in NAT hosts. NAT hosts will be infected during the worm propagation in the first layer at discrete time. The third layer shows worm propagation in inner hosts after the corresponding NAT hosts infected. The infected hosts in three layers interact with each other. That is an infected hosts in the first layer can infect hosts in the second layer. But it cannot infect hosts in the third layer. The infected hosts in the second layer can infect hosts in the first layer and hosts in the corresponding third layer. The infected hosts in the third layer can infect hosts in the three layers. The interaction of the three layers of TLWM is

shown in Fig. 3.

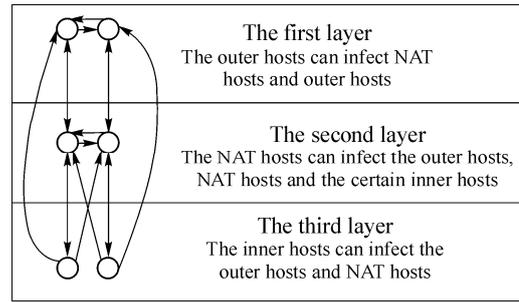


Fig. 3 The three layers of TLWM

We define notations added in TLWM model as follows:

- 1) β : the infection rate of the hosts in the first layer.
- 2) β' : the infection rate of the hosts in the second layer.
- 3) β'' : the infection rate of the hosts in the third layer.
- 4) p : the probability a NAT host is infected.
- 5) $I_n(t)$: the number of infected NAT hosts at time t .
- 6) $I_k(t)$: the number of infected inner hosts behind the k NAT at time t .
- 7) N_k : the number of hosts behind each NAT host.

The definitions of each notation are given as follows:

$$\beta = \frac{\eta}{2^{32}} \tag{7}$$

$$\beta' = p \frac{\eta}{2^{32}} \tag{8}$$

$$\beta'' = p \frac{\eta}{2^{32}} \tag{9}$$

$$\frac{dI(t)}{dt} = \left[\beta I(t) + \beta' I_n(t) + \sum_j \beta'' I_j(t) \right] [N - I(t)] \tag{10}$$

$$I = I(t) + I_n(t) + \sum_j I_j(t) \tag{11}$$

We consider infection process of worm in NAT hosts as a stochastic process. We model the worm propagation on the inner hosts by the simple worm model. Note that the beginning time of worm's outbreak in behind NAT is depending on if the NAT host is infected.

$$\frac{dI_k(t)}{dt} = \beta'' [I_k(t) + 1] [\Psi - I_k(t)] \tag{12}$$

According to the same principle, we also can derive corresponding equations from two-factor worm model. We use the same definition of β , β' , β'' . For the worm propagation on outer hosts, the process of modeling is as follows:

$$S(t + \Delta t) - S(t) = - \left[\beta I(t) + \beta' I_n(t) + \sum_j \beta'' I_j(t) \right] S(t) \Delta t - \frac{dQ(t)}{dt} \Delta t \Rightarrow$$

$$S(t+\Delta t) - S(t) = - \left[\beta I(t) + \beta' I_n(t) + \sum_j \beta'' I_j(t) \right] S(t) \Delta t - \frac{dQ(t)}{dt} \Delta t \quad \Rightarrow$$

$$\frac{dS(t)}{dt} = - \left[\beta(t) I(t) + \beta' I_n(t) + \sum_{j \neq k} \beta'' I_k(t) \right] S(t) - \frac{dQ(t)}{dt} \quad (13)$$

Note that $S(t) + I(t) + R(t) + Q(t) = N$ holds for any time t . Substituting $S(t) = N - I(t) - R(t) - Q(t)$ into Eq. (13)

yields the differential equation describing the behavior of the number of infectious hosts $I(t)$ as

$$\frac{dI(t)}{dt} = \left[\beta(t) I(t) + \beta' I_n(t) + \sum_{j \neq k} \beta'' I_k(t) \right] [N - S(t) - R(t) - Q(t)] - \frac{dR(t)}{dt} \quad (14)$$

The model of worm propagation on inner hosts can simply use the TFM. So the TLWM based on TFM is:

$$\left. \begin{aligned} \frac{dI(t)}{dt} &= \left[\beta(t) I(t) + \beta' I_n(t) + \sum_{j \neq k} \beta'' I_k(t) \right] [N - S(t) - R(t) - Q(t)] - \frac{dR(t)}{dt} \\ \frac{dI_j(t)}{dt} &= \beta''(t) [\Psi - R_j(t) - I_j(t) - Q_j(t)] I_j(t) - \frac{dR_j(t)}{dt} \\ \frac{dR(t)}{dt} &= \gamma I(t) \\ \frac{dQ(t)}{dt} &= \mu S(t) J(t) \\ \beta &= \frac{\eta}{2^{32}} \left[1 - \frac{I(t)}{N} + \Psi \right]^\eta \\ \beta' &= p \frac{\eta}{2^{32}} \left[1 - \frac{I_j(t)}{N} + \Psi \right]^\eta \\ \beta'' &= p \frac{\eta}{2^{32}} \left[1 - \frac{I_j(t)}{N} + \Psi \right]^\eta \\ I &= I(t) + I_n(t) + \sum_j I_j(t) \end{aligned} \right\} \quad (15)$$

5 Simulations

In the simulation, we reflect the effects of NAT on worm propagation based on the model in Eq. (10). The model does not have analytical solution. We use Matlab simulink to simulate the model we proposed above.

In order to simplify the TLWM worm model, we assume that the number of hosts behind each NAT is the same. There are m NAT hosts in the system. Then the worm model in Eq. (10) can be written as follows:

$$\frac{dI(t)}{dt} = [\beta I(t) + \beta' I_n(t) + m \beta'' I_j(t)] [N - I(t)] \quad (16)$$

For parameters $N=1\,100\,000$, $I_0=1$, $\eta=350$, $\beta=8.149 \times 10^{-8}$, $\beta'=8.149 \times 10^{-13}$, $\beta''=8.149 \times 10^{-13}$, $m=1\,000$, we obtain the numerical solutions for TLWM model and plot it in Fig. 4.

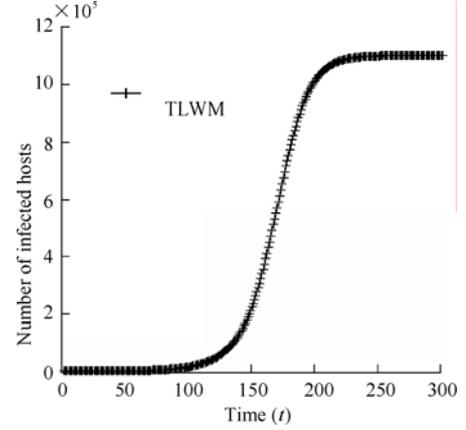


Fig. 4 Numerical solution of TLWM

In order to well understand the effects of NAT on worm propagation, the impact of the number of NAT is presented in Fig. 6. The dotted line is the worm propagation curve when m is 10. The solid line is when m is 100. So we can get that the more the value of m is, the faster the worm propagates. It is consistent with the conclusion in Ref. [13]. The Code Red v2 worm propagation curve is shown in Fig. 5 for comparison. We can see the TLWM model reflects the worm's propagation trend well, and it can well explain the effects of NAT on worm propagation quantitatively. What is the most important is it reflects the worm spread on a more real network environment accurately.

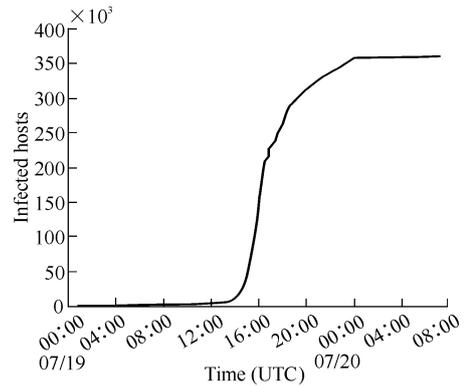


Fig. 5 The propagation of Code Red v2 worm

This model can well describe the worm propagation under NAT environment. According to adjust the value of m , we can see the effect of NAT on worm propagation is mainly on the ascending stage. The model proposed in Ref. [12] can only show the propagation trend like Fig. 4. It cannot demonstrate

worm propagation change when different number of NAT used in the network.

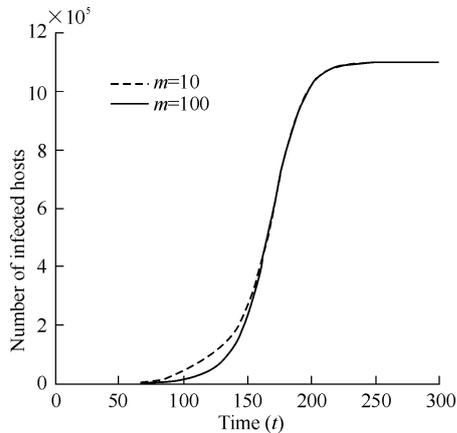


Fig. 6 Effect of NAT sizes on worm propagation

6 Conclusions

In this paper, we analyze the structure of Internet worm, and propose a three layer worm propagation model named TLWM to characterize the spread of the worm. This model represents the worm propagation under NAT environment. The first layer represents the hosts and routers on the Internet. The second layer is composed by the NAT hosts. The third layer is the hosts under each NAT. The infection rate between different layers is different. So the worm propagation under this environment is quite different from the homogeneous worm propagation.

The purpose of the TLWM is to get useful information about the effects of NAT on worm propagation. We can use the corresponding parameters to simulate other worms. Although previous worm model is more general, they do not take some network device into account. We model the TLWM with the same method as it is in TFM. In order to analyze the effects of NAT on worm propagation, we simulate it with different number of NAT hosts used in the system. The result shows that the use of NAT hosts will affect the beginning time of worm's outbreak.

Here, we only discuss the NATs which are homogeneous. With the development of new technologies, a majority of NATs are realized by NAT box which owns different OS and protocol stack. Therefore, as part of our ongoing work we are working on more complicated NAT environment and the development of effective defense techniques using the knowledge of worm propagation.

Acknowledgements

This work was supported by the Ministry of Education Science and Technology Basic Resource Data Platform (507001), the Ministry of Education Research Project for Returned Talents after Studying Abroad, and the Chinese Universities Scientific Fund (2009RC0502), the International Scientific and Technological Cooperation Program (S2010GR0902).

References

1. Feily M, Shahrestani A, Ramadass S. A survey of botnet and botnet detection. Proceedings of the 3rd International Conference on Emerging Security Information, Systems and Technologies (SECURWARE'09), Jun 18–23, 2009, Athens, Greece. Piscataway, NJ, USA: IEEE, 2009: 268–273
2. Li P, Salour M, Su X. A survey of internet worm detection and containment. IEEE Communications Surveys and Tutorials, 2008, 10(1): 20–35
3. Zou C C, Gong W B, Towsley D, et al. Code Red worm propagation modeling and analysis. Proceedings of the 9th ACM Conference on Computer and Communication Security (CCS'02), Nov 18–22, 2002, Washington DC, USA. New York, NY, USA: ACM, 2002: 138–147
4. Zou C C, Gong W B, Towsley D, et al. The monitoring and early detection of internet worms. IEEE/ACM Transaction on Networking, 2005, 13(5): 961–974
5. Chen Z, Gao L, Kwiat K. Modeling the spread of active worms. Proceedings of the IEEE 22nd Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM'03): Vol 3, Mar 30–Apr 3, 2003, San Francisco, CA, USA. Piscataway, NJ, USA: IEEE, 2003: 1890–1900
6. Su F, Lin Z W, Ma Y. Worm propagation modeling based on two-factor model. Proceedings of the 5th International Conference on Wireless Communications, Networking and Mobile Computing (WiCOM09), Sep 24–26, 2009, Beijing, China. Piscataway, NJ, USA: IEEE, 2009: 4p
7. Su F, Lin Z W, Ma Y. Effects of firewall on worm propagation. Proceedings of 2009 IEEE International Conference on Communications Technology and Applications (ICCTA'09), Oct 16–18, 2009, Beijing, China. Piscataway, NJ, USA: IEEE Computer Society, 2009: 880–884
8. Su F, Lin Z W, Ma Y. A survey of internet worm propagation models. Proceedings of the 2nd IEEE International Conference on Broadband Network & Multimedia Technology (IC-BNMT'09), Oct 18–20, 2009, Beijing, China. Piscataway, NJ, USA: IEEE, 2009: 453–457
9. Wang Y, Chakrabarti D, Wang C X, et al. Epidemic spreading in real networks: an eigenvalue viewpoint. Proceedings of the 22nd International Symposium on Reliable Distributed Systems (SRDS'03), Oct 6–18, 2003, Florence, Italy. Piscataway, NJ, USA: IEEE, 2003: 25–34
10. Okamura H, Kobayashi H, Dohi T. Markovian modeling and analysis of Internet worm propagation. Proceedings of the 16th IEEE International Symposium on Software Reliability Engineering (ISSRE'05), Nov 8–11, 2005, Chicago, IL, USA. Piscataway, NJ, USA: IEEE, 2005: 149–158
11. Kamra A, Feng H H, Misra V, et al. The effect of DNS delays on worm propagation in an IPv6 Internet. Proceedings of the 24th Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM'05): Vol 4, Mar 13–17, 2005, Miami, FL, USA. Piscataway, NJ, USA: IEEE, 2005: 2405–2414
12. Xing C Y, Yang L, Chen M. Modeling analysis of network worm propagation. Journal of University of Electronic Science and Technology of China, 2006, 36(3): 590–593 (in Chinese)
13. Rajab M A, Monroe F, Terzis A. On the impact of dynamic addressing on malware propagation. Proceedings of the 4th ACM Workshop on Recurring Malcode (WORM'06), Nov 3, 2006, Alexandria, VA, USA. New York, NY, USA: ACM, 2006: 51–56

(Editor: WANG Xu-ying)