Journal of Zhejiang University-SCIENCE C (Computers & Electronics)
ISSN 1869-1951 (Print); ISSN 1869-196X (Online)
www.zju.edu.cn/jzus; www.springerlink.com
E-mail: jzus@zju.edu.cn

Proactive worm propagation modeling and analysis in unstructured peer-to-peer networks*

Xiao-song ZHANG[1], Ting CHEN[†‡1], Jiong ZHENG[1], Hua LI[2]

([1]*School of Computer Science & Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China*)

([2]*Unit 78155 of People's Liberation Army, Chengdu 610016, China*)

[†]E-mail: chenting19870201@163.com

Received Mar. 14, 2009; Revision accepted Nov. 11, 2009; Crosschecked Oct. 31, 2009

Abstract: It is universally acknowledged by network security experts that proactive peer-to-peer (P2P) worms may soon engender serious threats to the Internet infrastructures. These latent threats stimulate activities of modeling and analysis of the proactive P2P worm propagation. Based on the classical two-factor model, in this paper, we propose a novel proactive worm propagation model in unstructured P2P networks (called the four-factor model) by considering four factors: (1) network topology, (2) countermeasures taken by Internet service providers (ISPs) and users, (3) configuration diversity of nodes in the P2P network, and (4) attack and defense strategies. Simulations and experiments show that proactive P2P worms can be slowed down by two ways: improvement of the configuration diversity of the P2P network and using powerful rules to reinforce the most connected nodes from being compromised. The four-factor model provides a better description and prediction of the proactive P2P worm propagation.

Key words: Proactive peer-to-peer (P2P) worm propagation modeling, Network topology, Configuration diversity, Attack and defense strategies, Four-factor model

doi:10.1631/jzus.C0910488 **Document code:** A **CLC number:** TP309.5; TP393.08

1 Introduction

A peer-to-peer (P2P) worm is a malicious code that makes use of a P2P network to spread from one machine to another (Khiat *et al.*, 2006). P2P systems become vehicles of worm propagation because of three intrinsic properties (Zhou *et al.*, 2005; Khiat *et al.*, 2006). Firstly, the large number of homogeneous P2P clients draws hackers' attention because users have to run the P2P clients in their terminals in order to acquire P2P services. Secondly, P2P topology accelerates worm propagation because worms become more efficient in searching for targets referring to neighbors' information. Thirdly, P2P worms are more arduous to detect and constrain because they do not make appreciable agitation to normal traffic (Yu *et al.*, 2006; Xia *et al.*, 2007).

In principle, we cast all P2P worms into three categories: passive P2P worms, reactive P2P worms and proactive P2P worms, each different in the pattern of propagation. Passive P2P worms copy themselves into the share folder of the P2P client and allure other users to download these copies and then complete propagation by running them in the peers' terminals (Thommes and Coates, 2006). Apparently, passive P2P worms cannot infect others without users' intervention. On the contrary, reactive and proactive P2P worms automatically propagate through common vulnerabilities of P2P clients (Chen and Gray, 2006). Reactive P2P worms infect only peers which are requesting files at that time while proactive P2P worms aim at infecting all vulnerable nodes as quickly as possible leveraging the cached neighbors' information

‡ Corresponding author

* Project (No. 09511501600) partially supported by the Science and Technology Commission of Shanghai Municipality, China

(Chen and Gray, 2006; Li *et al.*, 2009). Put simply, a proactive P2P worm is a more severe and fatal threat than passive and reactive P2P worms. Although there are currently only passive P2P worms in the wild (Singer, 2002; F-secure, 2004), security experts allege that proactive P2P worms will prevail in the near future because of the disclosure of the P2P clients' vulnerabilities according to the reports from antivirus organizations (Random Nut, 2003). Stimulated by the above-mentioned reasons, we are absorbed in the modeling and analysis of the proactive P2P worm propagation.

From a topological aspect, P2P networks belong to two categories: structured and unstructured. For the structured P2P network, the topology degree is actually a constant, while for the unstructured P2P network, the topology degree is a variable and represents a scale-free property (Ripeanu and Foster, 2002; Silvey and Hurwitz, 2004). Both of the two categovies are commonplace in practice. Content-addressable network (CAN) (Ratnasamy *et al.*, 2001), Chord (Stoica *et al.*, 2001), Pastry (Rowstron and Druschel, 2001) and Tapestry (Zhao *et al.*, 2004), for example, are structured P2P networks while Freenet and Gnutella (Adamic *et al.*, 2001; Ripeanu and Foster, 2002) are unstructured. This paper deals only with issues related to unstructured P2P networks while the investigation of structured P2P networks is a future consideration.

In this paper, we propose a novel proactive worm propagation model called the four-factor model because this model is an extension of the classical two-factor model in the unstructured P2P networks. One factor of the two-factor model is the dynamic countermeasures taken by Internet service providers (ISPs) and users, and the other is the slowed-down worm infection rate resulting from rampant propagation of worms causing congestion and trouble for some routers (Zou *et al.*, 2002). The first factor mentioned above is referred to in this paper and assimilated into the four-factor model, but we do not take the router congestion and trouble into consideration because proactive P2P worms just attack neighbors and do not generate excessive traffic.

Besides the users' countermeasures, we find three other factors affecting proactive P2P worm propagation. The first is the unstructured P2P topology, because the cached neighbors' information speeds up worm propagation. The second is the configuration diversity of nodes in the P2P network. That is, even though a worm breaks into a victim, it cannot infect other peers on condition that it does not execute properly in the alienated configuration. For example, a Windows worm becomes harmless when existing in Linux. Simulations show that improvement of the configuration diversity of the P2P network can apparently slow down the propagation of proactive P2P worms.

The last factor of the four-factor model is the attack and defense strategies. Although domination of all stages of worm propagation is unpractical, attackers have the ability to choose an elaborate strategy to speed up worm propagation. Experiments show that worms propagate faster when starting from the most connected node than from a random node. Although collaboration of antivirus software, firewall, intrusion detection systems and other security techniques can alleviate worm threats to a large extent, most users would not like to sacrifice too much performance for the sake of security. Thus, a deployable defense strategy should not impair the performance of the P2P networks appreciably. Simulations show that we can allay the threats of proactive P2P worms by protecting a few of the most connected nodes beforehand.

In this paper, we implement a simulation system based on the open source P2P simulation platform PeerSim (Sourceforge, 2009) and use the generalized linear preference (GLP) generator (Bu and Towsley, 2002) to generate scale-free topologies to represent P2P networks. Based on this simulation system, we make comparisons of the worm propagation trend with different parameters, and then validate the authenticity of our model. There are three main contributions of this work. Firstly, we present four factors affecting the proactive P2P worm propagation and propose the four-factor model. Secondly, we implement a simulation system for the simulation and analysis of proactive P2P worms. Thirdly, we find that improvement of the configuration diversity of P2P networks and guarding the most connected nodes against compromised can obviously slow down worm propagation. Worm prediction, damage assessment, detection and constraint can benefit from the four-factor model for its good profiling of the proactive P2P worm propagation.

2 Related works

2.1 Traditional models

Traditional models are those early models that do not take the P2P network topology into consideration. However, the bulk of the models in P2P networks are based on these traditional models. In the simple epidemic model (SEM), each host stays in one of two states: susceptible or infected. The state of any host can transit only from susceptible to infected (Zou *et al.*, 2003), so that in a finite network all hosts are infected in the end. While in the susceptible-infected-susceptible (SIS) model, each susceptible node can be infected and each infected node can be recovered and become susceptible again. Each host of the SIS model runs repeatedly through the cycle: susceptible→infected→susceptible (Wang and Wang, 2003).

The susceptible-infected-removed (SIR) model considers the removal process of the infected hosts. Each infected host of the SIR model has a probability of recovery or death. Once a host is removed from the infected host state, it will be immunized from that kind of worm and stay in the removed state forever. Each host of the SIR model either has the state transition susceptible→infected→removed or stays in the susceptible state forever (Frauenthal, 1980). The two-factor model is an improvement of the SIR model. It indicates that users' countermeasures and routers congestion can slow down worm propagation. Each host of the two-factor model either has the state transition susceptible→infected→removed or susceptible →removed so that in a finite network, all hosts are removed in the end (Zou *et al.*, 2002).

2.2 Evolutional models in P2P networks

The majority of the propagation models of proactive P2P worms are evolved from the traditional models as described in Section 2.1, so in this paper, we use the term 'evolutional models'. Some researchers (Yu *et al.*, 2008; Zhang *et al.*, 2008) transplanted the SEM model to P2P networks. They found that the P2P topology obviously accelerates proactive worm propagation since worms become more efficient in searching for targets referring to neighbors' information. Feng *et al.* (2008) proposed a model which is an improvement of the SIS model. There are two conclusions from their work: proactive P2P worms are difficult to constrain because of a rapid propagation speed, and worms propagate faster when starting from the most connected node than from the random or the least connected node.

3 Background of the two-factor model

The two-factor model proposed by Zou *et al.* (2002) indicates that there are two factors that most obviously impact worm propagation. One factor is the dynamic countermeasures taken by users such as cleaning, patching, and filtering. These countermeasures can transmit information about both susceptible and infected hosts to be removed. Let $R(t)$ denote the number of removed hosts from the infected hosts. The change $R(t)$ in unit time follows

$$\mathrm{d}R(t)/\mathrm{d}t = \gamma I(t), \tag{1}$$

where $I(t)$ denotes the number of infected hosts at time t; γ is the rate of removal of infected hosts and it is a constant. $Q(t)$ denotes the number of removed hosts from the susceptible hosts. The change of $Q(t)$ in unit time follows

$$\mathrm{d}Q(t)/\mathrm{d}t = \mu S(t)J(t), \tag{2}$$

where $S(t)$ is the number of susceptible hosts at time t; μ is the rate of removal of susceptible hosts and it is a constant; $J(t)$ denotes the number of infected hosts at time t and it follows the equation:

$$J(t) = I(t) + R(t). \tag{3}$$

The second factor is the congestion and trouble of Internet routers, which decreases the infection rate $\beta(t)$ (Cowie *et al.*, 2001; Wang *et al.*, 2002). Internet worms usually search for targets by scanning and many of those target IP addresses are never or rarely seen by routers when these routers work under normal conditions. Thus when the Internet is flooded with worms, the huge number of abnormal scanning packets can cause routers congested or reboot then slow down the worm propagation. The two-factor model models the decreased infection rate $\beta(t)$ by

$$\beta(t) = \beta_0 \left[1 - I(t)/N\right]^{\eta}, \tag{4}$$

where β_0 is the initial infection rate; the exponent η is

used to adjust the infection rate sensitivity to the number of infected hosts $I(t)$ and η=0 means a constant infection rate; N denotes the total number of hosts under consideration.

4 A new model: the four-factor model

It is difficult to model the propagation of proactive worms in unstructured P2P networks for three reasons. Firstly, the worm propagation process is complicate. Proactive P2P worms are not necessarily to infect other hosts immediately but can be activated by certain events, and thus the actual propagation process is discontinuous. Moreover, the time taken by proactive worms to infect other hosts is a variable and it is calculated on the worm size, distance, bandwidth and so on. Secondly, the P2P networks in practice are too huge to utilize. It is too expensive to deploy our worm monitors to the global P2P networks. Thirdly, there are not any known proactive P2P worms in the wild so we cannot examine our model by tracking worm propagation in real P2P networks.

In order to solve these three problems and obtain an analytical model, we provide simplifications and assumptions as follows. Firstly, we regard worm propagation as a continuous process. We assume that once a host is infected, it tries to infect all its neighbors immediately. We ignore the effect of worm size, distance, bandwidth, etc., and set the time duration for infecting other hosts to one unit time. Secondly, we generate scale-free topologies to represent P2P networks using the GLP generator (Bu and Towsley, 2002) (our experimental proactive P2P worms run in the generator). Thirdly, we implement experimental proactive P2P worms based on the open source P2P simulation platform PeerSim (Sourceforge, 2009).

4.1 Four factors affecting worm propagation

In terms of the propagation properties of proactive worms in the unstructured P2P networks, there are four factors which are not fully considered by former models:

1. Human countermeasures which result in removing both susceptible and infected hosts, as supposed in the classical two-factor model. We agree with Zou *et al.* (2002) and assimilate it into our four-factor model. When any user is aware of proactive P2P worms, the following actions can be taken to block worm propagation: cleaning compromised computers, patching or upgrading susceptible computers, setting up filters to block the worm traffic on firewalls or edge routers, or even disconnecting their computers from the Internet (Zou *et al.*, 2002).

2. P2P topology accelerates the propagation of proactive worms (Feng *et al.*, 2008; Yu *et al.*, 2008; Zhang *et al.*, 2008). Instead of randomly searching for victims, proactive P2P worms acquire the targets from the cached neighbors' information.

3. Configuration diversity (Zhou *et al.*, 2006; McIlwraith *et al.*, 2008) is capable of affecting worm propagation, but it is rarely considered in former models. Configuration diversity of each host in P2P networks can greatly decrease the overall vulnerability. Hosts with largely different configuration diversities are unlikely to be infected by the same worm (McIlwraith *et al.*, 2008). Configurations mentioned here include the operating system used, its version and patch level, additional software packages and executing applications with associated versions and open ports (McIlwraith *et al.*, 2008).

4. Attack and defense strategies can also impact the propagation of proactive P2P worms. We agree with Feng *et al.* (2008) that worms propagate faster when starting from the most connected node than from a random or the least connected node. We have a similar conclusion to Nie *et al.* (2008) that we can reduce worm propagation by immunizing a few of the most connected hosts prior to worm propagation. In this paper, we present two strategies: random and target. Random denotes that the process of selecting nodes is random while target denotes selecting the most connected nodes.

We do not take the routers' congestion into consideration since proactive P2P worms do not need to probe targets by blind scanning like Code Red worm (eEye Digital Security, 2001a; 2001b) and do not generate great traffic to flood routers.

4.2 Model description

Because our four-factor model is a general model, we cannot approach it with an analytical solution. Instead, we analyze the model in Section 5 based on numerical solutions of the difference equation using Matlab/Simulink (The Mathworks Inc., 2009).

Table 1 lists all notations used in this paper.

Table 1 Notations used in this paper

Notation	Description
N	Total number of hosts in the P2P network, $N=S(t)+I(t)+R(t)+Q(t)$
$S(t)$	Number of susceptible hosts at time t
$I(t)$	Number of infected hosts at time t
$R(t)$	Number of removed hosts from the infected population at time t
$Q(t)$	Number of removed hosts from the susceptible population at time t
$J(t)$	Number of hosts that are infected or recovered from infected population at time t, $J(t)=I(t)+R(t)$
γ	Rate of removal of infected hosts
μ	Rate of removal of susceptible hosts
$V(t)$	Set of newly infected hosts from susceptible population from time $t-1$ to t
d_i	Degree of host i
δ	Configuration diversity, $0\le\delta\le1$. The larger the δ, the larger the configuration diversity
ATTACK	Attack strategy, either random or target
DEFENSE	Defense strategy, either random or target

We assume that each host which is newly added in the infected population tries to attack all its neighbors immediately. If the new targets are already infected or removed, they will maintain their states. Otherwise, if the targets are in the susceptible population, they will be infected with a probability of $1-\delta$. Therefore, the larger the configuration diversity, the lower the infecting probability. Thus, if $\delta=0$, any susceptible host will be infected, while if $\delta=1$, none of the susceptible host will be infected. Enlightened by the work of Yu *et al.* (2008), the change in the number of infected hosts from time t to $t+1$ which are transmitted from susceptible hosts is

$$S(t)\times(1-\delta)\times\left[1-\left(1-\frac{1}{N}\right)^{\sum_{i\in V(t)}d_i}\right].$$

Proof Since the newly infected hosts will attack all their neighbors immediately, there are $\sum_{i\in V(t)}d_i$ attacks in the overall P2P networks in time t.

For any host in the P2P network, the probability of being attacked by one attack is $1/N$, and thus the probability of not being attacked is $1-1/N$. Then, for any host, the probability of not being attacked by $\sum_{i\in V(t)}d_i$ attacks is $\left(1-\frac{1}{N}\right)^{\sum_{i\in V(t)}d_i}$. So the probability of being attacked by at least one of $\sum_{i\in V(t)}d_i$ attacks is $1-\left(1-\frac{1}{N}\right)^{\sum_{i\in V(t)}d_i}$.

For any host, the probability of belonging to the susceptible population is $S(t)/N$. And if any susceptible host is attacked, it has a probability of $1-\delta$ to be infected. So the change in the number of infected hosts from time t to $t+1$ which are transmitted from susceptible hosts is $\left[1-\left(1-\frac{1}{N}\right)^{\sum_{i\in V(t)}d_i}\right]\times N\times(1-\delta)\times S(t)/N$ and the simplified form is $S(t)(1-\delta)\times\left[1-\left(1-\frac{1}{N}\right)^{\sum_{i\in V(t)}d_i}\right]$.

The infected population is increased by the infecting process and decreased by the immunizing process. So the change in $I(t)$ from time t to $t+1$ follows

$$I(t+1)-I(t)=S(t)\times(1-\delta)\times\left[1-\left(1-\frac{1}{N}\right)^{\sum_{i\in V(t)}d_i}\right]-[R(t+1)-R(t)]. \quad (5)$$

The susceptible population is decreased by both the infecting process and the immunizing process. So the change in $S(t)$ from time t to $t+1$ follows

$$S(t+1)-S(t)=-S(t)\times(1-\delta)\times\left[1-\left(1-\frac{1}{N}\right)^{\sum_{i\in V(t)}d_i}\right]-\left[Q(t+1)-Q(t)\right]. \quad (6)$$

We agree, as in the SIR model (Frauenthal, 1980) and the two-factor model (Zou *et al.*, 2002), the change of $R(t)$ depends only on $I(t)$ and we give the discrete form of the change of the removed population which is removed from infected hosts from time t to $t+1$ in Eq. (7):

$$R(t+1)-R(t)=\gamma I(t). \quad (7)$$

The removal process of our four-factor model is similar to that of the classical two-factor model (Zou *et al.*, 2002). Thus, we do not present details of modeling the change of $R(t)$ and $Q(t)$. The discrete form of the change in the number of removed population from susceptible hosts from time t to t+1 follows

$$Q(t+1)-Q(t)=\mu S(t)J(t). \tag{8}$$

Then we write down the complete difference equations of the four-factor model:

$$\begin{cases} I(t+1)-I(t)=S(t)\times(1-\delta)\times\left[1-\left(1-\dfrac{1}{N}\right)^{\sum_{i\in V(t)}d_i}\right] \\ \qquad\qquad -[R(t+1)-R(t)], \\ S(t+1)-S(t)=-S(t)\times(1-\delta)\times\left[1-\left(1-\dfrac{1}{N}\right)^{\sum_{i\in V(t)}d_i}\right] \\ \qquad\qquad -[Q(t+1)-Q(t)], \\ R(t+1)-R(t)=\gamma I(t), \\ Q(t+1)-Q(t)=\mu S(t)J(t), \\ J(t)=I(t)+R(t), \\ N=I(t)+S(t)+R(t)+Q(t), \\ I(0)=I_0, R(0)=0, Q(0)=Q_0, S(0)=N-I_0-Q_0. \end{cases} \tag{9}$$

We set $R(0)$ to 0 since, at the beginning of worm propagation, $I(0)=I_0<<N$, and most users are unaware of the worm. Thus, the removal process from the infected population is negligible (Zou *et al.*, 2002). We set $Q(0)$ to Q_0 rather than 0 because some hosts can be made immune from worm attacks by several security techniques such as antivirus, firewall, and intrusion detection. So we need to regard those immunized hosts as having the removed state at the beginning of worm propagation.

5 Numerical simulation and analysis

5.1 Simulation model

We use the tuple <N, I_0, Q_0, μ, γ, δ, ATTACK, DEFENCE> to represent the system parameters. We assume that in a default condition, there is only one host infected and none is immunized at the beginning of worm propagation. Then we set the initial value of N to 10 000, I_0 to 1 and Q_0 to 0. We assume that in the default condition once a susceptible host is attacked, it will be infected with probability 1, so the default value of δ is 0. We cannot obtain the actual value of μ or γ because there are no known proactive P2P worms in real P2P networks. So we set these two parameters according to the two-factor model: γ=0.05, μ=0.6/N as a default.

As a default, proactive worms select the initial victims at random. Since Q_0=0, we set the defense strategy as '/'. Thus the default parameter tuple of the four-factor model is <10 000, 1, 0, 0.6/N, 0.05, 0, random, />. All the following simulation results are the average of 100 simulations.

5.2 Numerical solutions of the four-factor model

We obtain the numerical solutions of the four-factor model and plot them in Fig. 1. Fig. 1 shows the behavior of $J(t)$=$I(t)$+$R(t)$, $I(t)$, and $Q(t)$ as functions of time t. We were impressed with the propagation speed of proactive worms and the size of the infected population: the time taken by worm propagation to reach the maximum of $I(t)$ is only 5 units of time and the maximum of $J(t)$ is approximately 0.8N.

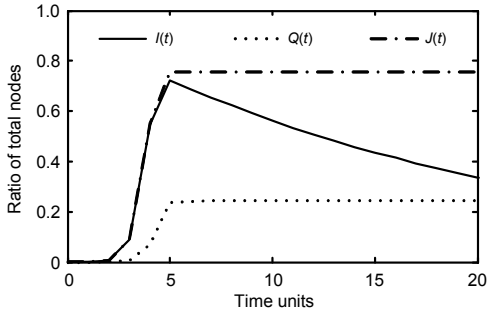

Fig. 1 Numerical solution of the four-factor model

Fig. 2 gives the comparison between the four-factor model and the two-factor model. We set all parameters of the two-factor model in accord with Zou *et al.* (2002) except that N=10 000. So the parameter tuple of our four-factor model is identical with that of the two-factor model. We obtain similar conclusions to others (Feng *et al.*, 2008; Yu *et al.*, 2008; Zhang *et al.*, 2008) that P2P topology accelerates worm propagation, and that it is more difficult to constrain proactive worms in a P2P network: the maximum of $J(t)$ of the four-factor model is about 30% larger than that of the two-factor model.

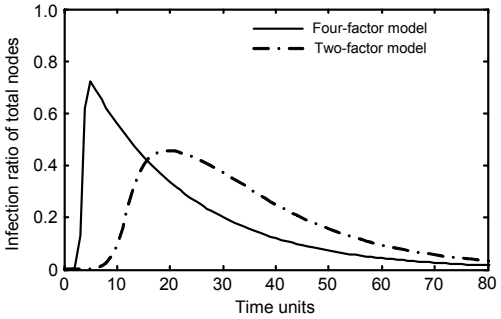

Fig. 2 Comparison of infection ratio between the four-factor model and the two-factor model

In this experiment, both the attack strategy and the defense strategy are set to random and the configuration diversity δ is 0, so the distinction of the four-factor model and the two-factor model in Fig. 2 is mostly due to the network topology. An unstructured P2P network is a kind of scale-free network while the two-factor model considers the network topology as homogeneous. The two-factor model assumes that an infected host is equally likely to infect any of other susceptible hosts and the hosts in the network can reach each other directly. Under this assumption, a great volume of worm traffic (most of the worm traffic is used to search for targets) will be generated so as to congest routers. In the scale-free network, however, an infected host can infect only its neighbors since any host can reach only its neighbors directly. As a result, without the congestion and trouble of routers, proactive worms can spread much quickly and infect more hosts in unstructured P2P networks than in homogeneous networks as supposed in the two-factor model.

5.3 Sensitivity of the system size

Fig. 3 shows the data on the attack performance of the four-factor model under different system sizes. We also present the data of the two-factor model for comparison in Fig. 3. The parameter tuple of the four-factor model in this experiment is <*, 1, 0, 0.6/*N*, 0.05, 0, random, /> and the system size is variable. The parameter tuple of the two-factor model is identical with that in Section 5.2 except the variable system size.

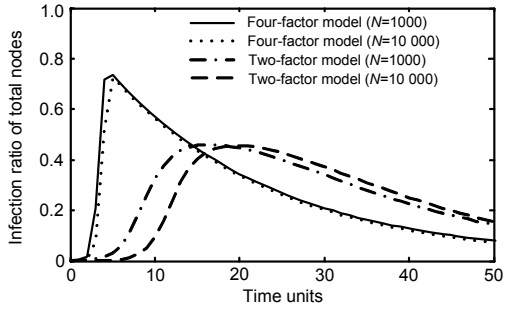

Fig. 3 Attack performance sensitivity of system size

We find that the attack performance of proactive worms in unstructured P2P networks is insensitive to the system size: the curve N=1000 and the curve N=10 000 are almost overlapped. While for the two-factor model, the attack performance declines a lot with the increasing system size because the process of searching for targets becomes more difficult. On the contrary, searching for a target is not a problem for proactive P2P worms no matter how large the system size because those worms need only to attack neighbors in the cache. Moreover, P2P networks of a larger size have more frequently connected nodes and these frequently connected nodes can accelerate worm propagation (Staniford *et al.*, 2002).

5.4 Sensitivity of the removal rate of the infected population

Fig. 4 depicts the data on the attack performance of the four-factor model under different γ. In this experiment, we set the parameter tuple as <10 000, 1, 0, 0.6/*N*, *, 0, random, /> and the rate of removal of infected hosts γ is a variable. We observe that the attack performance is insensitive to γ at the beginning of worm propagation (t≤5 units of time) but after the time when $I(t)$ reaches the maximum, $I(t)$ declines much faster with larger γ.

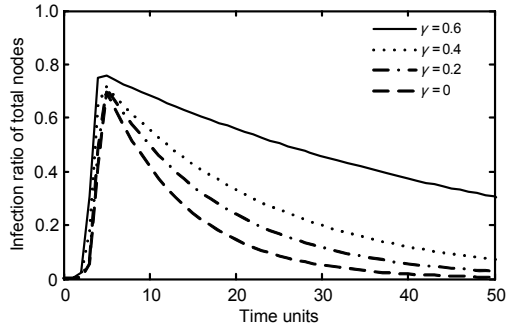

Fig. 4 Attack performance sensitivity to different rates of the infected population

The reason is that in the early stage of worm propagation, the process of infecting of susceptible hosts is much faster than that of removal of infected hosts, and thus the difference made by different γ is obscure, while in the middle and late stages, both the

process of infecting of susceptible hosts and removal of susceptible hosts cease.

5.5 Sensitivity of the removal rate of susceptible population

In this experiment, we set the parameter tuple as <10 000, 1, 0, *, 0.05, 0, random, /> and the rate of removal of susceptible hosts μ is a variable. Fig. 5 depicts the data on the attack performance under different μ. This experiment matches our expectation: the maximum of $I(t)$ declines acutely with the increase of μ for reasons that infected hosts are all transmitted from susceptible hosts and the process of removal of susceptible hosts becomes more rapid with larger μ.

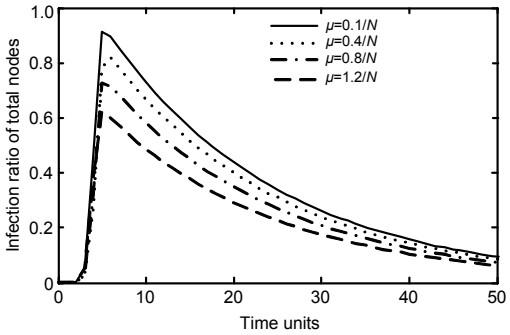

Fig. 5 Attack performance sensitivity to different rates of susceptible population

5.6 Sensitivity of the configuration diversity

In this experiment, we set the parameter tuple as <10 000, 1, 0, 0.6/N, 0.05, *, random, /> and the configuration diversity δ is a variable. Simulation results match our expectation: worm propagation gets slower and the number of infected hosts decreases greatly with increase in the configuration diversity. Fig. 6 validates our viewpoint that improvement of the configuration diversity is an efficient method to constrain proactive P2P worms.

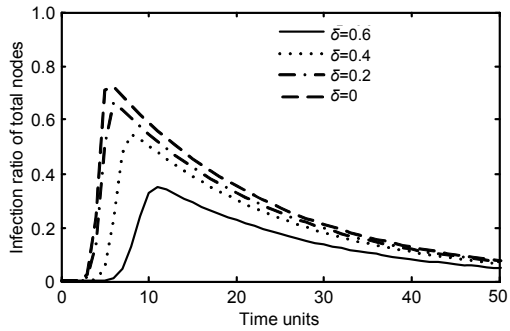

Fig. 6 Attack performance sensitivity to configuration diversity

5.7 Sensitivity of attack and defense strategies

All the experiments above assume that proactive worms select the initial victims at random and both the process of removal from susceptible and infected populations is also random. In the first experiment of this section, we set the initial victims as the most connected nodes and random nodes, and then we analyze the attack performance under different attack strategies. The parameter tuple of the experiment of attack strategies is <10 000, 1, 0, 0.6/N, 0.05, 0, *, />.

We obtain a similar conclusion to Feng *et al.* (2008) and Nie *et al.* (2008) that launching worms on frequently connected nodes causes faster worm propagation than on random initial nodes. The reason is that worm propagation starting from the most connected nodes will soon infect a lot of hosts and spread very quickly while that starting from the less frequently connected nodes has a low speed of propagation until enough nodes are infected. The simulation results in Fig. 7 validate our analysis.

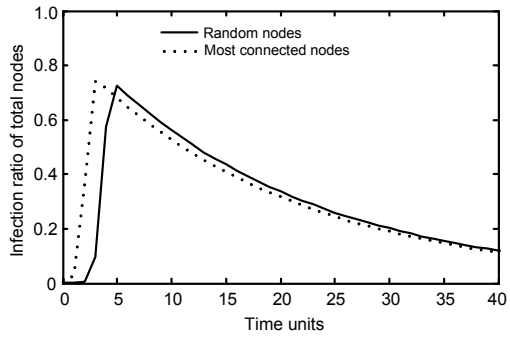

Fig. 7 Attack performance sensitivity to attack strategies

According to Cohen *et al*. (2000), the threshold of the fraction of nodes needed to be immunized when we immunize nodes randomly follows

$$p_c = 1 + \left(1 - m^{\alpha-2} M^{3-\alpha} \frac{\alpha-2}{3-\alpha}\right)^{-1}, \text{ if } 2 < \alpha < 3, \quad (10)$$

where m denotes the minimum of connections possessed by any host; M denotes the maximum; α represents the power law exponent. When we immunize nodes randomly, if the fraction of immunized hosts exceeds the threshold p_c, the network will be fragmented and worm propagation can be held back. In this paper, we set m=1, N=10 000 and α≈2.2 (which

is cited from the GLP generator), and then M is determined. We find that p_c calculated by Eq. (10) exceeds 0.9; that is to say, we need to immune more than 90% nodes randomly to cease worm propagation.

Fig. 8 shows the attack performance of proactive P2P worms under different defense strategies. We set the parameter tuple as <10 000, 1, *, 0.6/N, 0.05, 0, random, *> to examine our analysis above. In this experiment, the number of initially removed hosts Q_0 and the defense strategy are variables. From Fig. 8, we find that random immunization has little capacity to improve the global security: worms can infect almost all other nodes. On the contrary, worm propagation can be constrained by immunization of a few of the most connected nodes: about 20% nodes as shown in Fig. 8.

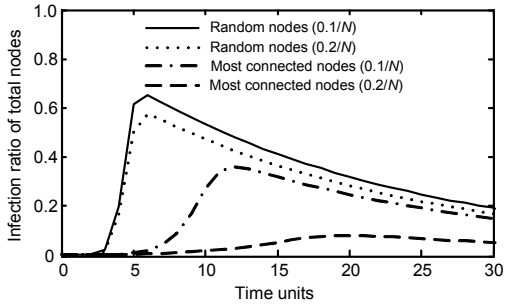

Fig. 8 Attack performance sensitivity to defense strategies

We use the theory of complex networks to explain the experiment in Fig. 8. There are many edges connected to the most connected nodes. If these nodes are removed, all edges connected to these nodes are also removed, and the connectivity of P2P networks decreases a lot. However, the global connectivity changes little when one removes a few of the less frequently connected nodes because they possess a small quantity of edges. We define the distance between each two directly connected nodes as one unit distance. Fig. 9 depicts the change of network diameters under random removing and target removing of the most connected nodes. Diameter is defined as the average length of the shortest paths between any two nodes in the network. Fig. 9 matches our expectations that unstructured P2P networks are robust to random failures, but vulnerable to target attacks of the most connected nodes (Albert *et al.*, 2000).

If the fraction of immunization of the most connected nodes exceeds a threshold p_s, networks will be fragmented. As p_s is much smaller than p_c, the number of nodes needed to cease worm propagation

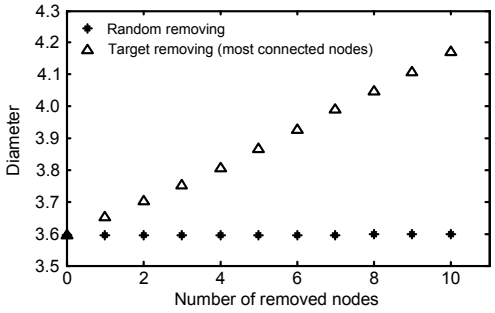

Fig. 9 Network diameters under different removing strategies (N=2000)

by immunizing the most connected nodes is far smaller than that of random immunization. Therefore, an effective method to constrain worm propagation is to use powerful rules to reinforce the most connected nodes from being compromised.

6 Discussion

The fact that the two-factor model matches the propagation of the Code Red worm quite well in the real world exemplifies that this model has an excellent potential to simulate worm behaviors in non-P2P networks. But in the condition of unstructured P2P networks, the four-factor model proposed in this paper prevails against the two-factor model. It is impossible to compare the two models directly in real P2P networks for lack of proactive P2P worms in practice. Nonetheless, we can deduce this conclusion from the following facts derived from former works as well as our four-factor model.

1. The propagation of proactive P2P worms in unstructured P2P networks is not sensitive to the change of the network size (an observation that contradicts the two-factor model).

2. Proactive P2P worms will not produce excessive traffic by attacking neighbors in the cache, and will not flood routers. But in the two-factor model the routers congestion factor is taken into consideration. Consequently, worm propagation of the two-factor model will be slower than that in practice.

3. The two-factor model cannot represent the fact that the propagation of proactive P2P worms can be slowed down by increasing the configuration diversity since it considers network topology as homogeneous: each host in the network is equal.

Unstructured P2P networks, however, are a kind of scale-free and inhomogeneous network.

4. The propagation of proactive P2P worms is affected by the attack/defense strategies. But this fact does not remain in the two-factor model because of the presumed equality of each host in that model.

7 Conclusion

In this paper, we propose a novel proactive worm propagation model: the four-factor model, which is an extension of the classical two-factor model to suit the unstructured P2P networks. The four factors are the human countermeasures which result in removing both susceptible and infected hosts, the P2P topology which results in accelerating worm propagation, the configuration diversity, and the attack and defense strategies. We do not consider the factor of routers congestion considered in the two-factor model because proactive P2P worms do not generate great traffic to flood routers.

We implement a simulation system based on PeerSim and a GLP network topology generator to examine the four-factor model. We obtain several conclusions from the numerical simulation in comparison with the two-factor model. Firstly, the unstructured P2P networks can speed up worm propagation and propagation of proactive P2P worms is not sensitive to the system size. Secondly, proactive P2P worms are difficult to constrain once the removing process lags behind the infecting process. Thirdly, worm propagation can be accelerated by launching worms on the frequently connected nodes rather than random nodes. Fourthly, one can constrain worm propagation by increasing the configuration diversity of the P2P network and protecting the most connected nodes from being compromised. The four-factor model is beneficial to the worm prediction, damage assessment, detection and constraint for its effective profiling of the proactive P2P worm propagation.

The four-factor model has its limitations. Firstly, it is suitable only for modeling continuously spreading worms and cannot predict the arbitrarily stopping or restarting worm events. Secondly, some parameters of this model are referenced upon the two-factor model because there are no known proactive P2P worms in the wild.

References

Adamic, L.A., Lukose, R.M., Puniyani, A.R., Huberman, B.A., 2001. Search in power-law networks. *Phys. Rev. E*, **64**(4): 461351-461358. [doi:10.1103/PhysRevE.64.046135]

Albert, R., Jeong, H., Barabdsi, A.L., 2000. Error and attack tolerance of complex networks. *Nature*, **406**(6794): 378-382. [doi:10.1038/35019019]

Bu, T., Towsley, D., 2002. On Distinguishing Between Internet Power Law Topology Generators. Proc. IEEE Conf. on Computer Communications, p.638-647. [doi:10.1109/INFCOM.2002.1019309]

Chen, G., Gray, R.S., 2006. Simulating Non-Scanning Worms on Peer-to-Peer Networks. Proc. 1st Int. Conf. on Scalable Information Systems, p.29-41. [doi:10.1145/1146847.1146876]

Cohen, R., Erez, K., Avraham, D.B., Havlin, S., 2000. Resilience of the Internet to random breakdowns. *Phys. Rev. Lett.*, **85**(21):4626-4628. [doi:10.1103/PhysRevLett.85.4626]

Cowie, J., Ogielski, A., Premore, B., Yuan, Y., 2001. Global Routing Instabilities During Code Red II and Nimda Worm Propagation. Available from http://www.renesys.com/projects/bgpinstability [Accessed on Aug. 8, 2002].

eEye Digital Security, 2001a. Analysis: .ida "Code Red" Worm. Available from http://www.eeye.com/html/Research/Advisories/AL20010717.html [Accessed on Mar. 22, 2008].

eEye Digital Security, 2001b. Analysis: Code Red II Worm. Available from http://www.eeye.com/html/Research/Advisories/AL20010804.html [Accessed on Sept. 12, 2005].

F-secure, 2004. Mydoom. Available from http//www.f-secure.com/tools [Accessed on Mar. 17, 2005].

Feng, C., Qin, Z., Cuthbet, L., Tokarchuk, L., 2008. Propagation Model of Active Worms in P2P Networks. Proc. 9th Int. Conf. for Young Computer Scientists, p.1908-1912. [doi:10.1109/ICYCS.2008.237]

Frauenthal, J.C., 1980. Mathematical Modeling in Epidemiology. Springer-Verlag, New York, p.1-7.

Khiat, N., Charlinet, Y., Agoulmine, N., 2006. The Emerging Threat of Peer-to-Peer Worms. Proc. 1st EEE Workshop on Monitoring, Attack Detection and Mitigation, p.1-3.

Li, Z., Zhang, Y., Hu, Z., Lin, H., Lu, C., 2009. Network-Based Detection Method Against Proactive P2P Worms Leveraging Application-Level Knowledge. Proc. 1st Int. Workshop on Education Technology and Computer Science, p.575-580. [doi:10.1109/ETCS.2009.661]

McIlwraith, D., Paquier, M., Kotsovinos, E., 2008. Di-Jest: Autonomic Neighbour Management for Worm Resilience in P2P Systems. Proc. IEEE Int. Symp. on a World of Wireless, Mobile and Multimedia Networks. [doi:10.1109/WOWMOM.2008.4594898]

Nie, X., Wang, Y., Jing, J., Liu, Q., 2008. Understanding the Impact of Overlay Topologies on Peer-to-Peer Worm Propagation. Proc. Int. Conf. on Computer Science and Software Engineering, p.863-867. [doi:10.1109/CSSE.2008.610]

Random Nut, 2003. The PACKET 0' DEATH FastTrack Network Vulnerability. Available from http://archive.cert.uni-stuttgart.de/bugtraq/2003/05/msg00277.html [Accessed on June 18, 2005].

Ratnasamy, S., Francis, P., Handley, M., Karp, R., Shenker, S., 2001. A scalable content addressable network. *ACM SIGCOMM Comput. Commun. Rev.*, **31**(4):161-172. [doi:10.1145/964723.383072]

Ripeanu, M., Foster, I., 2002. Mapping the gnutella network: macroscopic properties of large-scale peer-to-peer systems. *LNCS*, **2429**:85-93. [doi:10.1007/3-540-45748-8]

Rowstron, A., Druschel, P., 2001. Pastry: scalable, decentralized object location and routing for large-scale peer-to-peer systems. *LNCS*, **2218**:329-350. [doi:10.1007/3-540-45518-3]

Silvey, P., Hurwitz, L., 2004. Adapting Peer-to-Peer Topologies to Improve System Performance. Proc. Hawaii Int. Conf. on System Sciences, p.3117-3126.

Singer, M., 2002. Benjamin Worm Plagues KaZaA. Available from http://www.internetnews.com/bus-news/article.php/3531_1141841 [Accessed on Nov. 3, 2008].

Sourceforge, 2009. PeerSim P2P Simulator. Available from http://peersim.sourceforge.net/ [Accessed on Nov. 3, 2008].

Staniford, S., Paxson, V., Weaver, N., 2002. How to Own the Internet in Your Spare Time. Proc. 11th USENIX Security Symp., p.149-167.

Stoica, I., Morris, R., Karger, D., Kaashoek, M.F., Balakrishnan, H., 2001. Chord: a scalable peer-to-peer lookup service for Internet applications. *ACM SIGCOMM Comput. Commun. Rev.*, **31**(4):149-160. [doi:10.1145/964723.383071]

Thommes, R., Coates, M., 2006. Epidemiological Modeling of Peer-to-Peer Viruses and Pollution. Proc. 25th IEEE Int. Conf. on Computer Communications, p.181-192.

Wang, L., Zhao, X., Pei, D., Bush, R., Massey, D., Mankin, A., Wu, S., Zhang, L., 2002. Observation and Analysis of BGP Behavior under Stress. Proc. 2nd ACM SIGCOMM Workshop on Internet Measurment, p.183-195. [doi:10.1145/637201.637231]

Wang, Y., Wang, C., 2003. Modeling the Effects of Timing Parameters on Virus Propagation. ACM Workshop on Rapid Malcode, p.61-66. [doi:10.1145/948187.948198]

Xia, C., Shi, Y., Li, X., Gao, W., 2007. P2P worm detection based on application identification. *Front. Comput. Sci. China*, **1**(1):114-122. [doi:10.1007/s11704-007-0010-7]

Yu, W., Chellappan, S., Wang, X., Xuan, D., 2006. On Defending Peer-to-Peer System-Based Proactive Worm Attacks. Proc. IEEE Global Telecommunications Conf., p.1757-1761.

Yu, W., Chellappan, S., Wang, X., Xuan, D., 2008. Peer-to-peer system-based active worm attacks: modeling, analysis and defense. *Comput. Commun.*, **31**(17):4005-4017. [doi:10.1016/j.comcom.2008.08.008]

Zhang, Y., Li, Z., Hu, Z., Huang, Q., Lu, C., 2008. Evolutionary Proactive P2P Worm: Propagation Modeling and Simulation. Proc. 2nd Int. Conf. on Genetic and Evolutionary Computing, p.261-264. [doi:10.1109/WGEC.2008.75]

Zhao, B., Huang, L., Stribling, J., Rhea, S.C., Joseph, A.D., Kubiatowicz, J., 2004. Tapestry: a resilient global-scale overlay for service deployment. *IEEE J. Sel. Areas Commun.*, **22**(1):41-53. [doi:10.1109/JSAC.2003.818784]

Zhou, L., Zhang, L., McSherry, F., Immorlica, N., Costa, M., Chien, S., 2005. A First Look at Peer-to-Peer Worms: Threats and Defenses. Proc. 4th Int. Workshop of Peer-to-Peer Systems, p.24-35.

Zhou, Y., Wu, Z., Wang, H., Zhong, J., Feng, Y., Zhu, Z., 2006. Breaking Monocultures in P2P Networks for Worm Prevention. Proc. Int. Conf. on Machine Learning and Cybernetics, p.2793-2798. [doi:10.1109/ICMLC.2006.259000]

Zou, C.C., Gong, W., Towsley, D., 2002. Code Red Worm Propagation Modeling and Analysis. Proc. 9th ACM Conf. on Computer and Communication Security, p.138-147. [doi:10.1145/586110.586130]

Zou, C.C., Towsley, D., Weibo, G., 2003. On the performance of Internet worm scanning strategies. *Perform. Eval.*, **63**(7):700-723. [doi:10.1016/j.peva.2005.07.032]