Combating Good Point Set Scanning-based Self-learning Worms by Using Predators

Fangwei, Wang[1], Yunkai Zhang[1], Honggang Guo[1], and Changguang Wang[2]
(Corresponding author: Fangwei Wang)

Network Center, Hebei Normal University[1]
No.20, South ErHuan Rd., YuHua District, Shijiazhuang, 050024, China
College of Information Technical, Hebei Normal University[2]
No.20, South ErHuan Rd., YuHua District, Shijiazhuang, 050024, China
(Email: fw_wang@hebtu.edu.cn)
(Received Nov. 22, 2011; revised and accepted Apr. 24, 2012)

Abstract

Good point set scanning-based self learning worms can reach a stupendous propagation speed in virtue of the non-uniform vulnerable-host distribution. In order to terminate such self-learning worms, this paper proposes an interaction model. Using the interaction model, we obtain the basic reproduction number. The impact of different parameters of predators is studied. Simulations results show that the performance of our proposed model is effective in combating such worms, in terms of decreasing the number of hosts infected by the prey and reducing the prey propagation speed.

Keywords: network security, predator, interaction model, self-learning worms, equilibrium

1 Introduction

Internet worms can reduplicate themselves and attack computers which have vulnerability and are connected to the Internet without any human intervention. They have addressed a serious threat to confidentiality, integrity, and availability of computer resources on the Internet. Internet worms have reached a horrendous propagation speed because of their increasingly sophisticated spreading mechanisms. The time required for the infection of global targets has shrunk from days to minutes. Moreover, some worms exploiting the non-uniform vulnerable-host distribution may use advanced scanning strategies, e.g., good point set scanning, to infect a large number of hosts in a shorter time. This type of worms is called as self-learning worms. How to combat self-learning worms effectively is an urgent issue confronted by defenders.

The concept of predator is addressed by Toyoizumi [15]. The concept is to transform a malicious worm into a predator which spreads itself using the same mechanism as the original worm and immunizes a host. Some cases of predators, e.g., Welchia and CRClean worms, have been released to terminate Blaster and Code Red, respectively. However, they did not achieve the intended purpose, which create unprecedented dynamic and complex scenarios as well as detrimental effect on the Internet infrastructure [13].

Mathematical epidemiology is an important branch of science, aiming at devising optimal defense strategies to fight Internet worms. Several worm interaction models have been proposed [3, 5, 6, 10, 11, 12, 13, 14, 20], which are all based on epidemic models. However, those worm interaction models focusing on random-scan worm interactions have not considered the vulnerable-host distribution, are inadequate to model the war between self-learning worms and predators.

The goal of this paper is to mathematically model the behavior of containing worms. According to actual networks, we take the network-delay factor into account. This paper models prey-predator dynamics, further investigate the existence and stability of worm-free equilibrium point and endemic equilibrium point. We find that such effectiveness does not only depend on scan rate of predators but also on the number of groups.

The rest of the paper is organized as follows. Section 2 provides the background on vulnerable-host distributions, good point set scanning strategy, and predator/prey models. We propose a mathematical model in Section 3. Section 4 discusses the existence and stability of worm-free equilibrium point and endemic equilibrium point of the model. Section 5 studies the effect of some parameters on the infected population. We conclude our paper in Section 6.

2 Preliminaries

2.1 Distribution of Vulnerable Hosts

The distribution of vulnerable hosts in the Internet is not uniform. The reasons are as follows. Our measure results demonstrate that the vulnerable-host distribution is highly non-uniform by two collected data sets.

The first data set is a traffic log of the Witty worm obtained from CAIDA (the Cooperative Association for Internet Data Analysis) [1]. CAIDA used a Network Telescope approximately contains 2^{24} addresses. The collected data can accurately reflect the distribution of hosts which are vulnerable to the Witty worm [9]. The collected victim addresses are then formed a *group distribution* in 8 subnets, where

$$p_e(i) = \frac{number\ of\ addresses\ with\ the\ first\ byte\ equal\ to\ i}{total\ number\ of\ collected\ addresses}, \quad (1)$$

where $i = 0, 1, \cdots, 255$. The results are shown in Figure 1. It is observed that the distribution of vulnerable hosts is far from uniform.

The second data set is the web-server distribution. To estimate the distribution of web servers, we exploited two random uniform resource locator (URL) generators (http://www.uroulette.com and http://www.randomsite.net) to collect 46,082 random websites on October 16, 2007. Using a program written by the Perl language, however, we obtain 20,342 unique addresses of web servers. The results are shown in Figure 2. Although there are some differences from the values between Figure 1 and Figure 2, they demonstrate that the distributions of Witty-worm victims and web servers are far from uniform.

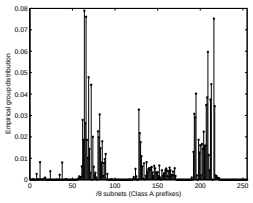

Figure 1: Uneven distribution of hosts infected by the Witty worm

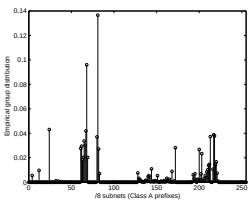

Figure 2: Uneven distribution of Web servers

2.2 Good Point Set Scanning

Good point set scanning (GPSS) [17] is inspired by the theories of the good point set in number-theoretic methods [4]. Due to a large IP address space and some IP addresses scanned many times, the probability of hitting a vulnerable host is very small, which results in a slow propagation speed and a small scale. The advantages of the GPSS are to reduce the number of scans needed for accurately and quickly attacking a large number of vulnerable hosts.

Let Z_n denote the IP address of an arbitrarily infected host, which consists of d ($d = 32$) bits and is seen as a d-dimensional cube. Each bit represents 1 or 0. That is, $Z_n = (z_1^n, z_2^n, \cdots, z_d^n)$. When using the GPSS to generate next IP address, the detail process is as follows.

In the d-dimensional unit cube, we set a good point set with n points.

$$P_n(i) = \{\{\alpha_1 \times i\}, \{\alpha_2 \times i\}, \cdots, \{\alpha_d \times i\}\}, \quad (2)$$

where $i = 1, 2, \cdots, n$, $\alpha_k = 2\cos(2\pi k/p)$, $1 \le k \le d$, p is the smallest prime number which satisfies $p \ge 2d+3$, $\{\nu\}$ denotes the fractional part of ν.

Among the n newly generated IP addresses, we assume that the k IP address $< B >^k$ is represented as $< B >^k = (b_1^k, b_2^k, \cdots, b_d^k)$, where $b_i^k = \{\alpha_i \times k\}$,

$$b_i^k = \begin{cases} 1 & if\ \{\alpha_i \times k\} \ge 0.5; \\ 0 & otherwise. \end{cases} \quad (3)$$

Using this method, we can obtain n newly generated IP addresses, which can be targeted by the worm. At each infected host, the worm generates newly IP addresses by the use of the good point set scanning and attacks the rest of vulnerable hosts.

2.3 Predator/Prey Models

In Reference [15], Toyoizumi et al. propose the use of predator as a defensive mechanism to protect the Internet from worms and viruses. The model employs the biologically inspired "Lotka-Volterra" equation to model the interaction of the predator-prey relationship between the malicious code and predator vaccination, with the goal of minimizing the number of predators required to eliminate the worm threat. The authors show that predators can be made to perform their tasks without flooding the network and consuming all available resources. Predators are benign programs that replicate and migrate from host to host across the Internet, which spread in much the same way malicious worms do but try to eliminate their designated "victim" worms.

Castaneda et al. [2] suggest modifying existing worms such as Code Red, Slammer and Blaster to terminate the original worm types, and discuss four anti-worm propagation schemes: a passive predator, an active scanning predator, an active-passive hybrid predator, and an intrusion detection system-based predator. The modified code will retain portion of attacking method so that it can choose and attack the same set of susceptible hosts. This paper assumes that the existence of this technology, and focuses on the first two predators. Active defense using beneficial worms is proposed [7]; however, the authors focus only on delay-limited worms with different type of interactions and do not consider network-related factors. In [10], Tamimi et al. propose some worm interaction

models based on Lotka-Volterra equations, do not take network-related factors into account. Tanachaiwiwat et al. investigate some worm interaction models focusing on random-scan worm interactions, and propose a new set of metrics to quantify effectiveness of one worm terminating other worm [11, 12, 13, 14]. The common problem of the above models is that they do not consider the non-uniform vulnerable-host distribution. Therefore, this paper emphasizes the effect of underlying network characteristics, e.g., reaction time, predator replication size, the number of groups on the interaction between self-learning worms and predators with the non-uniform vulnerable-host distribution.

3 Worm Interaction Model

The total population N is partitioned into four groups, and any host can potential be in any of these groups at any time tick t: *Susceptible (S)* - all hosts have not encountered the preys and have no circulating predators; *Prey (I_A)* - all hosts have encountered the preys in the Internet during the outbreak; *Predator (I_B)* - all hosts in this group are infected by predators and no longer susceptible to infection; *Recovered (R)*-all hosts in this group are not infectious and no longer susceptible to infection.

We base our model on the following assumptions: (1) We ignore the removal times. (2) The total population N is fixed, and does not vary with time t. (3) Once hosts are recovered or infected by predators, they have gained a certain period of permanent immunity and can no longer be infected by the same prey. This assumption is reasonable, because predators embedded relevant patches can guarantee hosts' security.

From the assumptions above, the standard incidence of the total population size can be expressed as

$$N = S(t) + I_A(t) + I_B(t) + R(t). \tag{4}$$

3.1 Infection Rate

A self-learning worm (named as a prey) replication can be significantly slowed down by network delay (ND) [13, 11, 12], e.g., transmission delay, link delay, processing delay, and queuing delay, which can momentously affect preys' propagation speed. Let φ_A and φ_B denote the network-delay factor which attenuates infection rate of prey and predator. Let s_A and s_B be the respective average scan rate of prey and predator. The Internet is partitioned into m groups. As shown in [9], the infection rate of prey β_A and the infection rate of predator β_B are

$$\begin{cases} \beta_A = \varphi_A s_A \sum_{i=1}^{m} \frac{p_g(i)p_g^*(i)}{\Omega_i}, \\ \beta_B = \varphi_B s_B \sum_{i=1}^{m} \frac{p_g(i)p_g^*(i)}{\Omega_i}, \end{cases} \tag{5}$$

where $p_g(i)$, referred to as the group distribution, is the percentage of live vulnerable hosts in group i ($i = 1, 2, \cdots, m$), Ω_i is the size of address space in group i, $p_g^*(i)$, referred to as the group scanning distribution, is the probability that a scan will hit group i.

Because self-learning worms (preys) can exploit the vulnerable host distribution, and scan the entire Internet according to this probability distribution, i.e., $p_g^*(i) = \frac{\sqrt{\Omega_i p_g(i)}}{\sum_{j=1}^{m}\sqrt{\Omega_j p_g(j)}}$, which is the optimal static strategy. If $\Omega_1 = \Omega_2 = \cdots = \Omega_m = \Omega/m$, $p_g^*(i) = \frac{\sqrt{p_g(i)}}{\sum_{j=1}^{m}\sqrt{p_g(j)}}$. The infection rate of prey and predator are

$$\begin{cases} \beta_A = \frac{s_A}{\Omega}\varphi_A \times m \sum_{i=1}^{m} \frac{\sqrt{p_g^3(i)}}{\sum_{j=1}^{m}\sqrt{p_g(i)}}, \\ \beta_B = \frac{s_B}{\Omega}\varphi_B \times m \sum_{i=1}^{m} \frac{\sqrt{p_g^3(i)}}{\sum_{j=1}^{m}\sqrt{p_g(i)}}. \end{cases} \tag{6}$$

Therefore, good point set scanning-based self-learning worms (preys) can increase the infection rate with the factor of $m \sum_{i=1}^{m} \frac{\sqrt{p_g^3(i)}}{\sum_{j=1}^{m}\sqrt{p_g(i)}}$, compared to random-scanning worms (where $\beta_A = \frac{s_A}{\Omega}\varphi_A$).

Let ND_A and ND_B denote the network delay for prey and predator, respectively. We can derive φ_A and φ_B as follows:

$$\begin{cases} \varphi_A = \frac{1}{1+s_A ND_A}, \\ \varphi_B = \frac{1}{1+s_B ND_B}. \end{cases} \tag{7}$$

The infection rate will be dynamic if the network congestion happens, which is consistent with the practical network. Let l be the number of targeted sub networks. For sub network i, let h_{Ai} and h_{Bi} denote the probability of network i being scanned for prey and predator, g_A and g_B be the worm replication size for prey and predator, q_{Ai} and q_{Bi} be the average queue length of outgoing links for prey and predator, bw_{Ai} and bw_{Bi} be the average bandwidth of outgoing links for prey and predator, c_{Ai} and c_{Bi} be the average packet drop rate for prey and predator, ld_{Ai} and ld_{Bi} be the average link delays for prey and predator. We can derive ND_A and ND_B as follows:

$$\begin{cases} ND_A = \sum_{i=1}^{l}(h_{Ai}(1-c_{Ai})(ld_{Ai} + \frac{g_A(q_{Ai}+1)}{bw_{Ai}})), \\ ND_B = \sum_{i=1}^{l}(h_{Bi}(1-c_{Bi})(ld_{Bi} + \frac{g_B(q_{Bi}+1)}{bw_{Bi}})). \end{cases} \tag{8}$$

3.2 Interaction Model

When there is a prey (A) and a predator (B), and the predator does not infect (or vaccinate) any susceptible host, but terminate any found prey, we consider this as infection-driven interaction. We propose the model represented in Figure 3 for the dynamics of the interaction in the Internet.

We assume that timeout period in this model is indefinite for both prey and predator. Let γ_S be the manual vaccination rate, and γ_A and γ_B be the manual removal rate for prey and predator, respectively. In our case γ_S is smaller than γ_A and γ_B. The influx of susceptible hosts comes from a constant recruitment Π. Let μ denote the death rate. from Figure 3, according to the theory of the

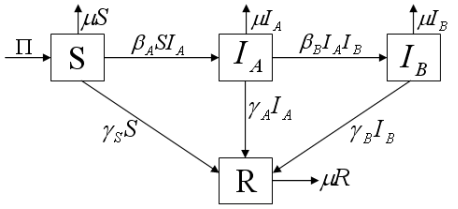

Figure 3: Interaction model

compartment model, we can easily write down the complete differential equations of the interaction Model (9).

$$\begin{cases} \frac{dS}{dt} = \Pi - \beta_A S I_A - \mu S - \gamma_S S, \\ \frac{dI_A}{dt} = \beta_A S I_A - \mu I_A - \beta_B I_A I_B - \gamma_A I_A, \\ \frac{dI_B}{dt} = \beta_B I_A I_B - \mu I_B - \gamma_B I_B, \\ \frac{dR}{dt} = \gamma_S S + \gamma_A I_A + \gamma_B I_B - \mu R. \end{cases} \tag{9}$$

4 Model Analysis and Basic Properties

Summing the equations in (9), we obtain that the total population N satisfies the differential equation

$$\frac{dN}{dt} = \Pi - \mu N. \tag{10}$$

Thus, we assume that the initial value is $N_0 = S_0 + I_{A0} + I_{B0} + R_0 = \frac{\Pi}{\mu}$ in order to have a population of constant size (that is, $S(t) + I_A(t) + I_B(t) + R(t) \equiv \frac{\Pi}{\mu}$). Obviously, the state variables $(S(t), I_A(t), I_B(t), R(t))$ remain in the biologically meaningful set $\Phi = \{(S, I_A, I_B, R) \in R_+^4 | 0 \le S + I_A + I_B + R \le \frac{\Pi}{\mu}\}$ for $(S(0), I_A(0), I_B(0), R(0)) \in R_+^4$, which is a positively invariant region.

Using $R(t) = \Pi/\mu - S(t) - I_A(t) - I_B(t)$ to eliminate $R(t)$ from the equations in (9) leads to the following reduced three-dimensional model:

$$\begin{cases} \frac{dS}{dt} = \Pi - \beta_A S I_A - \mu S - \gamma_S S, \\ \frac{dI_A}{dt} = \beta_A S I_A - \mu I_A - \beta_B I_A I_B - \gamma_A I_A, \\ \frac{dI_B}{dt} = \beta_B I_A I_B - \mu I_B - \gamma_B I_B. \end{cases} \tag{11}$$

The dynamical behavior of (9) on Φ is equivalent to that of (11). Thus, in the rest of the paper we will study Model (11) in the feasible region $\Phi_1 = \{(S, I_A, I_B) : S \ge 0, I_A \ge 0, I_B \ge 0, 0 \le S + I_A + I_B \le \frac{\Pi}{\mu}\}$ is also a positively invariant region for Model (11), and Model (11) is obviously well-posed in Φ_1.

Firstly, we derive the basic reproduction number of Model (11). It is easy to see that Model (11) always has a worm-free equilibrium, $P_0 = (S_0^*, I_{A0}^*, I_{B0}^*) = (\Pi/\mu, 0, 0)$.

Let $x = (I_A, I_B, S)^T$, then Model (11) can be written as

$$\frac{dx}{dt} = F(x) - Y(x),$$

where

$$F(x) = \begin{pmatrix} \beta_A S I_A \\ 0 \\ 0 \end{pmatrix},$$

$$Y(x) = \begin{pmatrix} \mu I_A + \beta_B I_A I_B + \gamma_A I_A \\ -\beta_B I_A I_B + \mu I_B + \gamma_B I_B \\ -\Pi + \beta_A S I_A + \mu S + \gamma_S S \end{pmatrix}.$$

Differentiating $F(x)$ and $V(x)$ with respect to I_A, I_B, S and evaluating at the worm-free equilibrium $P_0 = (S_0^*, I_{A0}^*, I_{B0}^*) = (\Pi/\mu, 0, 0)$, we have

$$F(P_0) = \begin{pmatrix} F_{2\times 2} & 0 \\ 0 & 0 \end{pmatrix},$$

$$Y(P_0) = \begin{pmatrix} & Y_{2\times 2} & 0 \\ & & 0 \\ \beta_A \Pi/\mu & 0 & \mu + \gamma_S \end{pmatrix},$$

where

$$F_{2\times 2} = \begin{pmatrix} \beta_A \Pi/\mu & 0 \\ 0 & 0 \end{pmatrix},$$

$$Y_{2\times 2} = \begin{pmatrix} \mu + \gamma_A & 0 \\ 0 & \mu + \gamma_B \end{pmatrix}.$$

$F(P_0)Y^{-1}(P_0)$ is the next generation matrix for Model (11). It then follows that the spectral radius (the largest absolute eigen value) of the matrix $F(P_0)Y^{-1}(P_0)$. Thus,

$$\rho(F(P_0)Y^{-1}(P_0)) = \frac{\beta_A \Pi}{\mu(\mu + \gamma_A)}.$$

According to Theorem 2 in [16], the basic reproduction number of Model (11) is

$$R_0 = \frac{\beta_A \Pi}{\mu(\mu + \gamma_A)}. \tag{12}$$

Let the right-hand side of equalities in Model (11) be zero, then calculating straightforwardly we can obtain $I_A = 0$ or $I_A > 0$.

For the case of $I_A = 0$, we have the worm-free equilibrium $P_0 = (S_0^*, I_{A0}^*, I_{B0}^*) = (\Pi/\mu, 0, 0)$.

For the case of $I_A > 0$, we can obtain the endemic equilibrium $P^*(S^*, I_A^*, I_B^*)$, where

$$S^* = \frac{\Pi}{\beta_A I_A^* + \mu + \gamma_S}, I_A^* = \frac{\mu + \gamma_B}{\beta_B},$$

$$I_B^* = \frac{\Pi - (\beta_A I_A^* + \mu + \gamma_S)(\mu + \gamma_A)}{\beta_B(\beta_A I_A^* + \mu + \gamma_S)}.$$

We have the following results on the stability of equilibrium P_0 and P^*:

Theorem 4.1. *The worm-free equilibrium P_0 is locally asymptotically stable if $R_0 < 1$, and unstable if $R_0 > 1$.*

Proof. According to $P_0 = (S_0^*, I_{A0}^*, I_{B0}^*) = (\Pi/\mu, 0, 0)$, the Jacobian matrix at the worm-free equilibrium P_0 is

$$J(P_0) = \begin{pmatrix} -\mu-\gamma_S & -\beta_A S_0^* & 0 \\ 0 & \beta_A S_0^* - \mu - \gamma_A & 0 \\ 0 & 0 & -\mu-\gamma_B \end{pmatrix}.$$

The corresponding eigenvalues of $J(P_0)$ are

$$\begin{cases} \lambda_1 = -\mu - \gamma_S, \\ \lambda_2 = \beta_A S_0^* - \mu - \gamma_A, \\ \lambda_3 = -\mu - \gamma_B. \end{cases} \quad (13)$$

All parameters of the model are assumed to be positive. Therefore, for λ_1, λ_3 to be negative, i.e., for a worm-free equilibrium to be locally asymptotically stable, the following condition has to be required: $(\beta_A\Pi - \mu^2 - \gamma_A\mu)/\mu < 0$. By the stability theory [8], the sufficient condition for the three-dimensional model to be asymptotically stable is that $\lambda_i < 0$, for $i = 1, 2, 3$. It is easy to show $\lambda_1 < 0$ and $\lambda_3 < 0$ in Model (11). As to $\lambda_2 < 0$ is equal to $S_0^* < \frac{\mu+\gamma_A}{\beta_A}$. If we substitute $S_0^* = \Pi/\mu$ into the above inequality, we have $\frac{\beta_A\Pi}{\mu(\mu+\gamma_A)} < 1$, which is exactly the sufficient condition in the lemma. □

Further, we can obtain the following theorem.

Theorem 4.2. *The worm-free equilibrium P_0 is globally asymptotically stable if $R_0 \leq 1$.*

Proof. Learn from the first equation of Model (11)

$$S^{'}(t) \leq \Pi - (\mu + \gamma_S)S(t).$$

Thus

$$S(t) \leq \frac{\Pi}{\mu} + (S(0) - \frac{\Pi}{\mu})exp[-\mu t],$$

when $t \to \infty$, we obtain $S(t) \leq \frac{\Pi}{\mu}$.

Let us consider the following Lyapunov function defined by

$$L(t) = I_A(t).$$

The time derivative of $L(t)$ along the solution of Model (11) is given by

$$\begin{aligned} L^{'}(t) &= I_A^{'}(t) \\ &= \beta_A S I_A - \mu I_A - \gamma_A I_A - \beta_B I_A I_B \\ &\leq \beta_A S I_A - \mu I_A - \gamma_A I_A \\ &\leq \beta_A I_A \Pi/\mu - (\mu + \gamma_A) I_A \\ &= I_A[\tfrac{\beta_A \Pi}{\mu} - (\mu + \gamma_A)] \\ &\leq 0. \end{aligned}$$

Thus, we prove that the worm-free equilibrium P_0 is globally stable. This completes the proof. □

For the case of the endemic equilibrium P^*. The Jacobian matrix at P^* is

$$J(P^*) = \begin{pmatrix} C & -\beta_A S^* & 0 \\ \beta_A I_A^* & D & -\mu - \gamma_B \\ 0 & \beta_B I_B^* & 0 \end{pmatrix},$$

where $C = -\mu - \gamma_S - \beta_A I_A^*$, and $D = \beta_A S^* - \mu - \gamma_A - \beta_B I_B^*$.

The eigenfunction of $J(P^*)$ is $f(\lambda) = \lambda^3 + a_1\lambda^2 + a_2\lambda + a_3$, where

$$\begin{aligned} a_1 &= \beta_B I_B^* + \gamma_A + \beta_A I_A^* + 2\mu + \gamma_S > 0, \\ a_2 &= \gamma_S\beta_B I_B^* + \mu^2 + \mu\gamma_S + 2\mu\beta_B I_B^* - \gamma_S\beta_A S^* \\ &+ \beta_A\beta_B I_A^* I_B^* + \gamma_S\gamma_A + \gamma_B\beta_B I_B^* + \gamma_A\beta_B I_A^* \\ &+ \mu\gamma_A + \beta_A I_A^*\mu - \beta_A\mu S^*, \\ a_3 &= \gamma_S\gamma_B\beta_B I_B^* + \beta_A\beta_B\gamma_B I_A^* I_B^* + \mu^2\beta_B I_B^* \\ &+ \beta_A\beta_B\mu I_A^* I_B^* + \mu\beta_B I_B^*\gamma_B + \gamma_S\mu\beta_B I_B^* > 0. \end{aligned}$$

By the Routh-Hurwitz theorem, the Routh-Hurwitz array for P^* is as follows:

$$\begin{pmatrix} 1 & a_2 \\ a_1 & a_3 \\ (a_1a_2 - a_3)/a_1 & 0 \\ a_3 & 0 \end{pmatrix}.$$

Thus, if we can verify that $(a_1a_2 - a_3)/a_1$ has the same sign with a_2, then the three eigenvalues all have negative real parts. Obviously, $a_1 > 0, a_3 > 0$. Also, $(a_1a_2 - a_3)/a_1 > 0 \Longleftrightarrow a_1a_2 - a_3 > 0$ holds by the little algebraic calculation of a_1, a_2, a_3. Thus, the Routh-Hurwitz stability conditions are satisfied, which implies that the endemic equilibrium P^* is locally asymptotically stable.

From the above discussion, we can summarize the following conclusion.

Theorem 4.3. *As long as $R_0 > 1$ holds, the endemic equilibrium P^* is locally asymptotically stable.*

Theorem 4.4. *The endemic equilibrium P^* is globally asymptotically stable if $R_0 > 1$.*

Proof. It is easy to see that the model has unique positive equilibrium P^* if $R_0 > 1$ holds. Then we consider the following Lyapunov function [18] defined as

$$L(t) = \int_{S^*}^{S} \frac{x - S^*}{x} dx + \int_{I_A^*}^{I_A} \frac{x - I_A^*}{x} dx. \quad (14)$$

The time derivative of $L(t)$ along the solution of Equation (11) is given by

$$\begin{aligned} L^{'}(t) &= (\tfrac{S-S^*}{S})S' + (\tfrac{I_A - I_A^*}{I_A})I_A^{'} \\ &= (1 - \tfrac{S^*}{S})[\Pi - \beta_A S I_A - \mu S - \gamma_S S] \\ &+ (1 - \tfrac{I_A^*}{I_A})[\beta_A S I_A - \mu I_A - \beta_B I_A I_B - \gamma_A I_A] \\ &\leq (1 - \tfrac{S^*}{S})[\Pi - \beta_A S I_A - \mu S - \gamma_S S] \\ &+ (1 - \tfrac{I_A^*}{I_A})[\beta_A S I_A - \mu I_A - \gamma_A I_A] \\ &= -\Pi(\tfrac{S}{S^*})(\tfrac{S^*}{S} - 1)^2 \\ &\leq 0. \end{aligned}$$

Thus, we prove that the endemic equilibrium P^* is globally stable. This completes the proof. □

5 Experiments

5.1 Simulation Settings

Our main goal is to verify the accuracy of our mathematical model and have better understanding of worm infection in a rich set of environments. We choose the Slammer-like self-learning worm as basic behavior of a prey in this experiment. Slammer worm is chosen because, despite its simplicity, it still holds the world record of fastest-spread worm yet [19]. The Blaster worm of 2003 infected at least 100,000 Microsoft Windows systems. Therefore, in our simulation, we assume that the total vulnerable population is $N = 100,000$. Slammer is a bandwidth-limited worm with an average scan rate $s_A = 4000$ scans/second [19]. A "bandwidth-limited worm" is a worm that fully uses the link bandwidth of an infected host to send out infection traffic. Slammer worm uses UDP scan to transfer worm replication to random chosen vulnerable hosts of the network. Each UDP infection packet sent out by Slammer is 404 bytes ($g_A = 404$) [19]. We also assume $I_A(0) = 10$, i.e., 10 vulnerable hosts in the system are infected by the worm at the beginning.

To implement Model (9), we must first obtain the group distribution $p_g(i)$. However, the group distribution is unknown before the Slammer-like worm is released. Therefore, we can use the web-server distribution as an example of real vulnerable-host distribution. The empirical distribution $p_e(i)$ (Equation (1)) can be used to reflect the relative distribution of the number of web servers as a function of the first byte values.

Thus, we assume

$$p_g(i) = p_e(i). \tag{15}$$

We simulate prey (A) and predator (B) which may have different scan rates, initial number of infected hosts and the same group distribution information. We assume that the average scan rate of predators is $s_B = 4000$ scans/second, the worm replication size of predators is $g_B = 404$ bytes, and the initial infective of predators is $I_B(0) = 1$.

In order to obtain the authentic network delay, we generate a two-level topology with 1000 vulnerable hosts. It can help us test our model with bottleneck network having large number of hops (1-4 hops) with moderate bandwidth (512 kbps, 10Mbps local network) and delay between hosts (1 ms on average). The topology has 10 local networks; each local networks has 100 hosts with one of them acting as a router. One AS (Autonomous System) has one or two local networks. We use BRITE Internet topology generator to generate the links between routers. We obtain the relatively actual network delay ($ND_A = ND_B = 0.011$ seconds).

Other parameters in these simulations are given as follows: the manual vaccination rate of susceptible hosts is $\gamma_S = 6 \times 10^{-6}$; manual removal rates for prey and predator are $\gamma_A = 6 \times 10^{-4}$ and $\gamma_B = 5 \times 10^{-4}$, respectively. The death rate is $\mu = 0.00001$. The results are based on the average of at least 10 simulation runs.

5.2 Performance Evaluations

The basic reproduction number is $R_0 = 0.294$ through the calculation by the use of the above parameter values. The prey will gradually disappear from the theory. From Figure 4, we can clearly see that the tendency of the prey propagation is depressive, which is consistent with the theory analysis. Hosts infected by the prey vanish and the network, in the long term, is in a good operational state. Finally, all vulnerable hosts are vaccinated or recovered, and become healthy hosts that are no longer infected by the prey.

Intuitively, predators must have a greater scan rate in order to combat preys effectively. To study the effect of scan rate of predators, we run the simulation with $s_B = 100, 1000, 2000, 4000, 8000$ scans/second and show the simulation results in Figure 5. When we increase the scan rate of predators, the number of hosts infected by preys should have an obvious decrease from the theory. However, Figure 5 does not represent our expectation. Predators can combat preys effectively even though they have a relatively small scan rate. Thus, the scan rate of predators plays an unimportant role in combating preys.

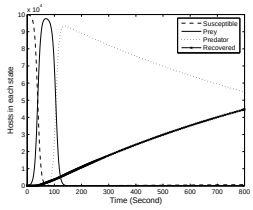

Figure 4: Dynamics with an asymptotically stable worm-free equilibrium point

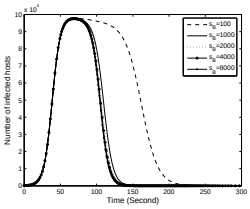

Figure 5: The effects of scan rate of predators

Figure 6 shows the effect of changing the network delays (which vary between 5ms and 22ms) on prey propagation. As expected, a larger network delay results in diminishing the prey propagation speed, lowering the total number of infected hosts, and prolonging the time at which infected population reaches its peak. Network delay ND_A and ND_B rely mainly on transmission delay, link delay, processing delay, and queuing delay, which can momentously affect preys' propagation speed.

Predator replication size g_B is a transmission overhead reflecting the efficiency of coding and compression technique that automatic generation or programmer uses. Figure 7 shows the effect of changing the predator replication size (which vary between 404 and 1212 bytes) on

prey propagation. To our surprise, the increase of predator replication size almost has no impact on the number of hosts infected by preys, and has a slightly decrease on the propagation speed of preys.

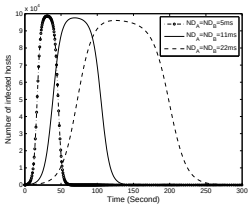

Figure 6: The effects of network delay

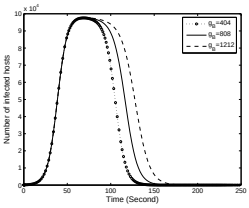

Figure 7: The effects of replication size of predators

The reaction time is the time required before releasing predators by automated worm-generation or programmer. In general, it takes some time (T) from the appearance of preys to the generation of predators. That is, in Model (9), each $I_B(t)$ should be replaced by the corresponding $I_B(t-T)$. Our interaction models assume that predators have been generated in advance, with emphasis on the effect of the reaction time. Our model do not take the delay of patch applying into account. Furthermore, we also assume that patch takes effect instantly without need of rebooting the host. Figure 8 shows the effect of changing the reaction time (which vary between 10 and 40 seconds) on prey propagation. From Figure 8, the conclusion can be drawn that the longer the reaction time is, the longer the prey infection prolongs. Also we can see that just because of the effect of predators, the prey infection descends rapidly after predators burst out. The reaction time is a key parameter that plays an important role in combating preys effectively.

What happens if we enlarge the number of groups m? To study the effect of the number of groups, we run the simulation with $m = 256, 65536$ and show the simulation results in Figure 9. From Figure 9, we can see that a larger m in the infection-driven and vulnerability-driven models can more effectively combat preys, which can terminate preys in a shorter time. The number of groups (m) play an important role in combating preys.

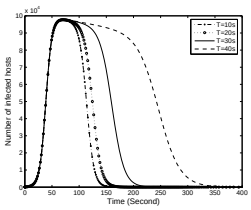

Figure 8: The effects of reaction time of predators

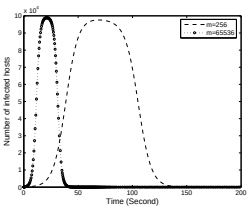

Figure 9: Effect of the number of groups

6 Conclusions and Future Directions

This paper proposed an interaction model based on based on Lotka-Volterra equations and deduced the conditions for the existence and stability of the worm-free equilibrium and endemic equilibrium point for the interaction model. Simulation results showed the propagation of preys being mainly governed by the network delay, the number of groups and the reaction time of predators. However, the scan rate and replication size of predators did not significantly affect the number of hosts infected by the prey and the preys' propagation speed. This can provide an important guideline in the control of self-learning worms.

Our future work will validate the model with simulations obtained by NS2 (Network Simulator Version 2), and compare our model with existing models to verify the accuracy. We will also focus on the following key techniques: the control strategies of predators; the optimal time of releasing predators; detecting methods of self-learning worms; the traceability and safety of predators.

Acknowledgments

This research was supported by the Natural Science Foundation of Hebei Province of China under No. F2009000322, Foundation of Hebei Normal University under No. L2010B21, the Applied Basic Research Programs of Science and Technology Commission Foundation of Tianjin No. 09JCYBJC00500 and Foundation of JiaCheng.

References

[1] K. Brian, M. Jim, and W. Pat. "Ucsd network telescope – witty worm dataset,". http://www.caida.org/data/passive/witty_worm _dataset.xml.

[2] F. Castaneda, E. C. Sezer, and J. Xu, "Worm vs. worm: preliminary study of an active counter-attack mechanism," in *Proceedings of the 2004 ACM workshop on Rapid malcode*, pp. 83–93, Washington DC, USA, October 2004.

[3] X. Han and Q. Tan, "Dynamical behavior of computer virus on internet," *Applied Mathematics and Computation*, vol. 217, no. 6, pp. 2520–2526, 2010.

[4] J. G. Liao, "Variance reduction in gibbs sampler using quasi random numbers," *Journal of Computational and Graphical Statistics*, vol. 7, no. 3, pp. 253–266, 1998.

[5] B. K. Mishra and G. M. Ansari, "Differential epidemic model of virus and worms in computer network," *Applied Mathematics and Computation*, vol. 14, no. 3, pp. 149–155, 2012.

[6] B. K. Mishra and S. K. Pandey, "Dynamic model of worms with vertical transmission in computer network," *Applied Mathematics and Computation*, vol. 217, no. 21, pp. 8438–8446, 2011.

[7] D. Nicol and M. Liljenstam, "Models and analysis of active worm defense," in *proceeding of Mathematical Methods, Models and Architecture for Computer Networks Security Workshop*, pp. 38–53, Petersburg, Russia, September 2005.

[8] R. C. Robinson, *An introduction to dynamical system: continuous and discrete)*. USA: Prentice Hall, 2004.

[9] C. Shannon and D. Moore, "The spread of the witty worm," *IEEE Security & Privacy*, vol. 2, no. 4, pp. 46–50, 2004.

[10] Z. M. Tamimi, "Model-based analysis of two fighting worms," in *IEEE/IIU Proceedings of International Conference on Computer & Communication Engineering (ICCCE'06)*, pp. 157–163, kuala Lumpur, Malaysia, May 2006.

[11] S. Tanachaiwiwat and A. Helmy, "Analyzing the interactions of self-propagating codes in multi-hop networks," in *The eighth International Symposium on Stabilization, Safety, and Security of Distributed Systems (SSS 2006)*, pp. 582–583, Dallas, Texas, USA, November 2006.

[12] S. Tanachaiwiwat and A. Helmy, "Computer worm ecology in encounter-based networks," in *Fourth International Conference on Broadband Communications, Networks and Systems (BROADNETS 2007)*, pp. 535–543, Raleigh, North Carolina, USA, September 2007.

[13] S. Tanachaiwiwat and A. Helmy, "Modeling and analysis of worm interactions (war of the worms)," in *Fourth International Conference on Broadband Communications, Networks and Systems (BROADNETS 2007)*, pp. 649–658, Raleigh, North Carolina, USA, September 2007.

[14] S. Tanachaiwiwat and A. Helmy, "Encounter-based worms: Analysis and defense," *Ad Hoc Networks*, vol. 7, no. 7, pp. 1414–1430, 2009.

[15] H. Toyoizumi and A. Kara, "Predators: Good will mobile codes combat against computer viruses," in *Proceedings of the 2002 workshop on New security paradigms*, pp. 11–17, Virginia Beach, VA, USA, September 2002.

[16] P. Van den Driessche and J. Watmough, "Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission," *Mathematical Biosciences*, vol. 180, no. 1, pp. 29–48, 2002.

[17] F. Wang, Y. Zhang, and J. Ma, "Modeling and analysis of a self-learning worm based on good point set scanning," *Wireless Communications and Mobile Computing*, vol. 9, no. 4, pp. 573–586, 2009.

[18] W. Xu and Z. Zhang, "Global stability of sir epidemiological model with vaccinal immunity and bilinear incidence rates," *College Mathematies*, vol. 19, no. 6, pp. 76–80, 2003.

[19] W. Xu and Z. Zhang, "Inside the slammer worm," *IEEE Security & Privacy*, vol. 1, no. 4, pp. 33–39, 2003.

[20] F. Yang, H. Duan, and X. Li, "Modeling and analyzing of the interaction between worms and antiworms during network worm propagation," *Science in China Series F: Information Sciences*, vol. 48, no. 1, pp. 91–106, 2005.

Fangwei Wang received his B.S. degree in 2000 from College of Mathematics & Information Sciences, Hebei Normal University, his M.S. degree in 2003 from College of Computer Science and Software, Hebei University of Technologyis, his Ph.D degree in 2009 from College of Computer at Xidian University. Currently he is an associate professor at Hebei Normal University, Shijiazhuang, China. His research interests include: network and information security, sensor networks.

Yunkai Zhang received his B.S. degree in 1986 from Department of Electronic and Information Engineering, Hebei University, his M.S. degree in 1997 from Department of Telecommunication Engineering, Beijing University of Posts and Telecommunications, and his Ph.D degree in 2005 from College of Computer at Xidian University. Currently he is a professor at Hebei Normal University, Shijiazhuang, China. His research interests include network and information security.

Honggang Guo received his B.S. degree in 2000 from College of Mathematics & Information Sciences, Hebei Normal University, his M.S. degree in 2004 from College of Electrical & Electronic Engineering, Huazhong University of Science & Technology. His research interests include: network and information security.

Changguang Wang received his M.S. degree in 1996 from School of Physical Science and Technology, Sichuan University, and his Ph.D degree in 2009 from College of Computer at Xidian University. Currently he is an associate professor at Hebei Normal University, Shijiazhuang, China. His research interests include network and information security.