A Closed-Form Expression for Static Worm-Scanning Strategies

Zesheng Chen
Department of Electrical & Computer Engineering
Florida International University
Miami, FL 33174
zchen@fiu.edu

Chao Chen
Department of Engineering
Indiana University - Purdue University Fort Wayne
Fort Wayne, IN 46805
chen@engr.ipfw.edu

***Abstract*—This work presents a closed-form expression for characterizing the spread of static worm-scanning strategies through a mean-field approximation. Our model can both accurately capture the worm propagation speed before the number of infections becomes large and explicitly demonstrate the effects of important parameters such as the vulnerable-host distribution and the worm-scanning strategy. Our approach is based on the mean-field theory that investigates the average number of infected hosts over time. Experiments are carried out based on the parameters chosen from the Witty worm. Experimental results verify that the closed-form expression can accurately reflect the mean value of infections over time before the infected hosts become saturated for a wide range of scanning methods. Therefore, our model can help defenders design better detection and defense systems and provide a stepping stone towards obtaining closed-form expressions for the propagation of dynamic worm-scanning strategies.**

I. Introduction

Worm attacks present a significant threat to the Internet. A worm can self-propagate across the Internet in a short time by exploiting security flaws on vulnerable hosts without human intervention. Thus, worms, such as Code Red, Slammer, and Witty, have infected hundreds of thousands of hosts and caused enormous damages. Most worms use a scanning technique that selects a target in an IP address space and then sends out a probe to attempt to compromise this target. Among all scanning methods, *random scanning* is the simplest method that selects a target at random in the IPv4 address space and has been widely used by real worms. Recent studies, however, have shown that worms can potentially apply more advanced scanning strategies, such as *hitlist scanning* [11], *routable scanning* [13], [15], *importance scanning* [4], [3], and *OPT-STATIC* [12]. These advanced scanning strategies have been demonstrated to be able to spread a worm much faster than the random-scanning method. Therefore, it is imperative that defenders would model the spreading behaviors of advanced worm-scanning strategies accurately.

Vojnovic *et al.* pointed out that studying worm-scanning methods is also of interest in a wide variety of areas such as streaming broadcasting, database maintenance, and Web-service membership management [12]. These applications potentially exploit epidemic-style information dissemination techniques to spread information among participants quickly. Therefore, modeling epidemic-style information dissemination or worm-scanning strategies can provide further understandings to these areas.

Most advanced worm-scanning strategies take advantage of the non-uniform distribution of vulnerable hosts over subnets. For example, *importance scanning* probes the Internet according to an underlying vulnerable-host distribution and forces worm scans on the most relevant parts of an address space [4]. For these advanced scanning strategies, the Internet is grouped into subnets according to such standards as the IP prefix, autonomous systems, and the first byte of IP addresses (/8 subnets). Since the distribution of vulnerable hosts over subnets has been observed to be highly uneven [10], [12], [5], [6], a worm would spend more scans on subnets that contain many vulnerable hosts to speed up worm propagation. That is, a worm scans different subnets with different likelihoods so that a subnet containing more vulnerable hosts would be hit by a worm scan with a higher probability.

We call the probabilities of scanning different subnets as the *group scanning distribution*. If the group scanning distribution is fixed at all times, a subnet would be hit by a worm scan with a fixed probability, and such strategies are named *static worm-scanning strategies*, including random scanning, static importance scanning [3], and OPT-STATIC [12]. Otherwise, the group scanning distribution varies with time, and such strategies are called *dynamic worm-scanning strategies*, including localized scanning [8], [1], dynamic importance scanning [3], and K-FAIL [12]. This work focuses on static worm-scanning strategies. Specifically, we attempt to derive a closed-form expression for the spread of static strategies. Hopefully, our approach can provide a stepping stone towards finding closed-form expressions for the propagation of dynamic strategies, which are more complex and difficult to obtain.

Several approaches have been proposed to model the spread of worms. *Stochastic* models have been studied to capture the variance of worm propagation at the early stage [9], [7]. Stochastic models, however, may require extensive computations and focus only on the early stage of worm spreading. Instead, most analytical models of worm propagation have used *deterministic* dynamic equations, ignoring the variance of worm infection [11], [16], [8], [2]. For example, the analytical active worm propagation (AAWP) model forms a dynamic equation to characterize the expected number of infected hosts

over time [2]. Based on dynamic equations, however, it is difficult to understand the effects of important parameters (*e.g.*, the vulnerable-host distribution and the group scanning distribution) on worm propagation. Moreover, except for extreme cases (*e.g.*, random scanning), it is nearly impossible to derive an exact closed-form expression from the dynamic equation. An alternate approach to both model worm propagation and capture parameters' effects has been proposed by Vojnovic *et al.* [12]. The authors formulate worm infection as an *optimization* problem and focus on the number of worm scans required to reach a predetermined fraction of vulnerable hosts. In Section II, however, we point out that two worm-scanning strategies can use the same number of worm scans to infect the same number of hosts, but differ significantly in the worm propagation speed. Therefore, to characterize both the worm propagation speed and the parameters' effects, we derive a *closed-form expression* from the deterministic dynamic equation through a *mean-field* approximation in Section III.

A mean-field approach provides a way to gain some insight into the behavior of complex systems at a relatively low cost [17]. Specifically, the mean-field method focuses on the averages of the system, ignoring fluctuations. In this work, we neglect the fluctuation of the number of infected hosts and derive the average of the infections in each subnet. We further apply the Taylor expansion and focus on the first-order term. In this way, we obtain a closed-form expression for the spread of static worm-scanning strategies. In Section IV, our experiments show that the closed-form expression can characterize the worm propagation speed before the infected hosts become very large (even beyond the early stage). Moreover, our closed-form expression explicitly demonstrates the effects of the vulnerable-host distribution and the group scanning distribution.

Characterizing the worm propagation speed before infections become large is a key element to worm detection and defense [14]. If a worm can compromise a large number of hosts before it is detected, it is too late for defenders to slow down the worm. Therefore, it is critical that defenders would detect and fight against a worm before it infects too many hosts. Thus, our closed-form expression provides an accurate picture for defenders to understand the average of the worm spreading speed in the time window of detection and defense.

The remainder of this paper is structured as follows. Section II motivates this work. Section III derives a closed-form expression for the spread of static worm-scanning strategies. Section IV further verifies our model through experiments. Section V concludes this paper.

II. Motivations

In [12], worm propagation is formulated as an optimization problem: minimizing the number of worm scans required to reach a predetermined fraction of vulnerable hosts. The authors designed the optimal static strategy and proved that the performance of such an optimal static strategy achieves that of the optimal dynamic strategy in terms of the minimum number of worm scans. The authors, however, ignored the dynamic behaviors of worm spreading. That is, although two strategies use the same number of worm scans to infect the same number of vulnerable hosts, the propagation speeds can be very different. To show this, we apply the extension of the AAWP model [2] (Equation (21) in [4]) to characterize the spread of the optimal static strategy (Equation (10) in [12]) and the optimal dynamic strategy (Equation (24) in [4]). Figure 1 shows the spreading behaviors of these two strategies using the /8 subnets distribution of Witty-worm victims [10]. Both strategies use 1.76×10^9 worm scans to infect 90% of vulnerable hosts. However, while the optimal static strategy uses 102 seconds to infect 90% vulnerable hosts, the optimal dynamic strategy takes only 56 seconds. Therefore, even though both the number of worm scans (or samplings) and the number of infected hosts for these two strategies are equal, the worm dynamic behaviors differ significantly. This motivates us to approach the problem of modeling worm propagation from a different aspect: to derive a closed-form expression for the time required for a worm to infect a certain number of vulnerable hosts.

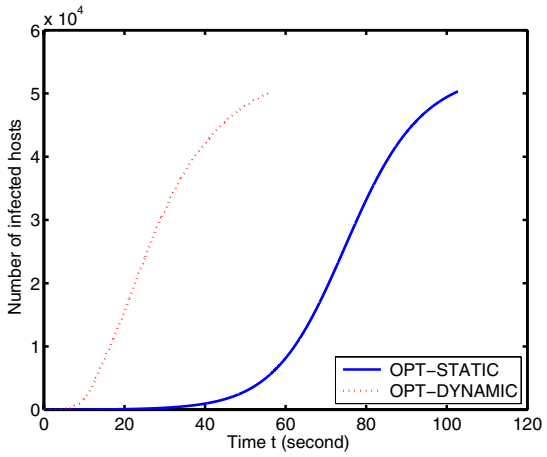


Fig. 1. Compare the optimal static strategy with the optimal dynamic strategy.

III. A Closed-Form Expression for Static Worm-Scanning Strategies

In this section, we derive a closed-form expression for modeling the spread of static worm-scanning strategies through a mean-field approximation.

A. Static Worm-Scanning Strategies

Assume that the Internet contains Ω IP addresses (*i.e.*, $\Omega = 2^{32}$) and totally N vulnerable hosts. Here, the Internet is assumed to be divided into m groups. Group i ($i = 1, 2, \cdots, m$) contains Ω_i IP addresses and has totally N_i vulnerable hosts, where $\sum_{i=1}^{m} \Omega_i = \Omega$ and $\sum_{i=1}^{m} N_i = N$. A worm scans group i with probability q_i, where $\sum_{i=1}^{m} q_i = 1$. Thus, q_i's form a group scanning distribution. In this paper, q_i's are fixed at all times, representing static worm-scanning methods. Let s be the worm scanning rate or the rate at which an infected host scans an address space for a vulnerable host. Suppose that

there are $S_i(t)$ uninfected vulnerable hosts and $I_i(t)$ infected hosts in group i at time t, where $S_i(t) + I_i(t) = N_i$. At time t, there are $sI(t)q_i$ worm scans hitting group i, where $I(t)$ is the total number of infected hosts in the Internet and $I(t) = \sum_{i=1}^{m} I_i(t)$. Thus, $I_i(t)$ can be characterized by a classic *susceptible* $\rightarrow$ *infected* (SI) epidemic model and follows a dynamic differential equation:

$$\frac{dI_i(t)}{dt} = sI(t)\frac{q_i S_i(t)}{\Omega_i}. \tag{1}$$

Summing up $i = 1, 2, \cdots, m$ and using $S_i(t) = N_i - I_i(t)$, we have

$$\frac{dI(t)}{dt} = sI(t)\left(\sum_{i=1}^{m}\frac{q_i N_i}{\Omega_i} - \sum_{i=1}^{m}\frac{q_i I_i(t)}{\Omega_i}\right). \tag{2}$$

Note that if a worm uses random scanning, *i.e.*, $q_i = \Omega_i/\Omega$, Equation (2) becomes a *logistic equation* [18] and can lead to a well-known closed-form solution [2]:

$$t = \frac{\Omega}{sN}\ln\frac{I(t)[N - I(0)]}{I(0)[N - I(t)]} \tag{3}$$

or

$$I(t) = \frac{I(0)N}{I(0) + [N - I(0)]e^{-sNt/\Omega}}. \tag{4}$$

In general, however, it is difficult to derive a closed-form expression of $I(t)$ based on Equation (2).

B. Mean-Field Approximation

To get a closed-form expression of $I(t)$, we define $u(t)$ as the total number of worm scans sent by all infected hosts by time t, *i.e.*,

$$u(t) = s\int_0^t I(x)dx. \tag{5}$$

Since there are $q_i u(t)$ scans that hit group i among $u(t)$ scans, the mean value of the number of infected hosts in group i can be derived by

$$I_i(t) = N_i\left[1 - \left(1 - \frac{1}{\Omega_i}\right)^{q_i u(t)}\right], \tag{6}$$

where $1 - \left(1 - \frac{1}{\Omega_i}\right)^{q_i u(t)}$ is the probability that a vulnerable host in group i is hit by at least one worm scan. The above equation applies a mean-field approach that neglects the fluctuation of the number of infections in group i and focuses on the average. We apply the Taylor expansion and get

$$I_i(t) = u(t)\frac{q_i N_i}{\Omega_i} + O\left(\frac{1}{\Omega_i^2}\right). \tag{7}$$

Assuming that $\Omega_i >> 1$ and $q_i u(t)$ is not very large, we can obtain the approximation of the average of the number of infected hosts in group i or the mean-field approximation:

$$I_i(t) \approx u(t)\frac{q_i N_i}{\Omega_i}. \tag{8}$$

Summing up $i = 1, 2, \cdots, m$, we have

$$u(t) = \frac{I(t)}{\sum_{i=1}^{m} q_i N_i/\Omega_i}. \tag{9}$$

Plugging Equations (8) into Equation (2), we have

$$\frac{dI(t)}{dt} = sI(t)\left(\sum_{i=1}^{m}\frac{q_i N_i}{\Omega_i} - u(t)\sum_{i=1}^{m}\frac{q_i^2 N_i}{\Omega_i^2}\right). \tag{10}$$

Applying Equation (9), Equation (10) becomes

$$\frac{dI(t)}{dt} = sI(t)\left(\sum_{i=1}^{m}\frac{q_i N_i}{\Omega_i} - \frac{\sum_{i=1}^{m} q_i^2 N_i/\Omega_i^2}{\sum_{i=1}^{m} q_i N_i/\Omega_i}I(t)\right). \tag{11}$$

The above equation is in the form of a *logistic equation* [18] and can lead to a closed-form expression. Set $A = s\sum_{i=1}^{m} q_i N_i/\Omega_i$ and $B = \frac{\left(\sum_{i=1}^{m} q_i N_i/\Omega_i\right)^2}{\sum_{i=1}^{m} q_i^2 N_i/\Omega_i^2}$. The solution for Equation (11) is

$$t = \frac{1}{A}\ln\frac{I(t)[B - I(0)]}{I(0)[B - I(t)]} \tag{12}$$

or

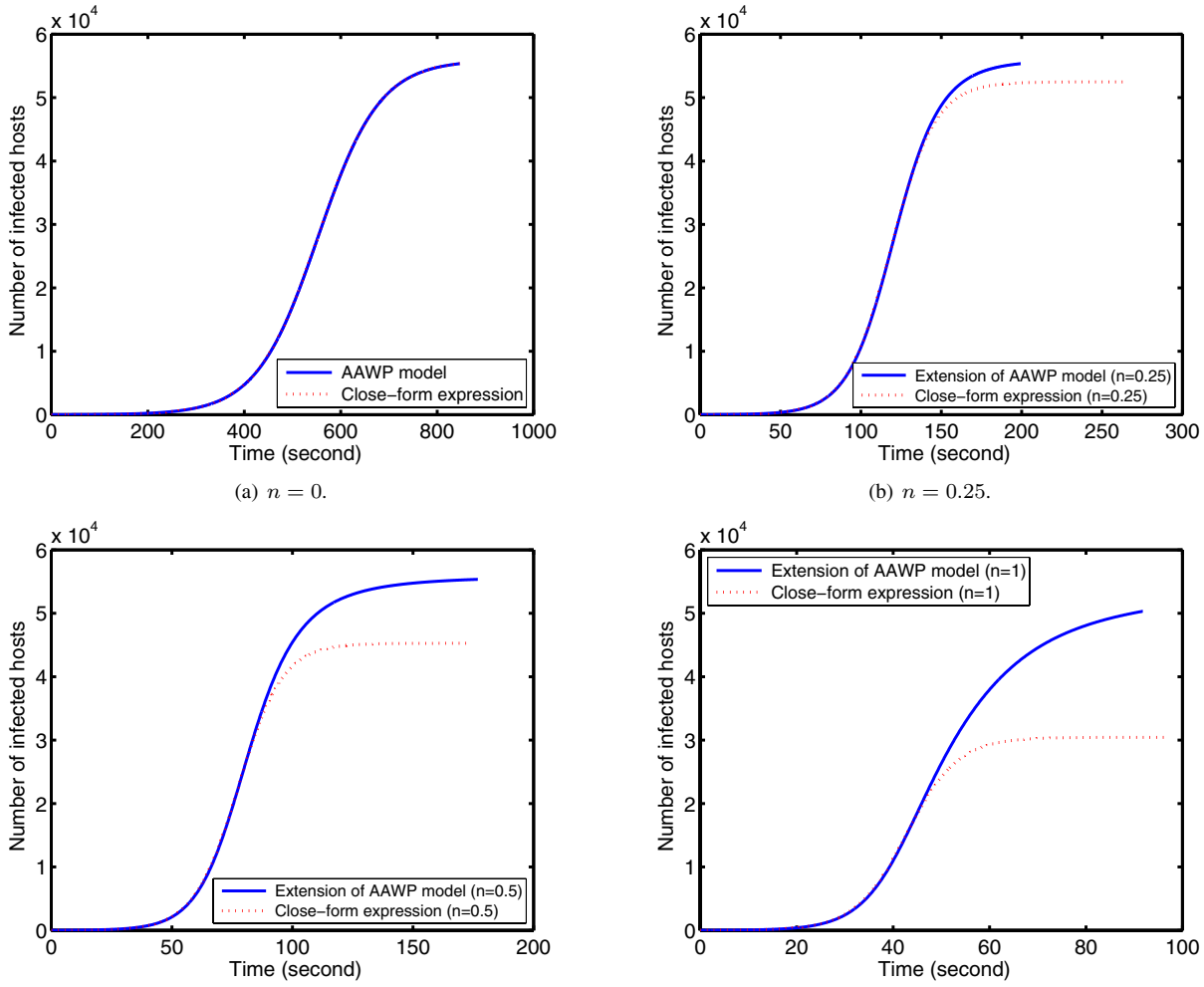
$$I(t) = \frac{I(0)B}{I(0) + [B - I(0)]e^{-At}}. \tag{13}$$

Note that if a worm uses random scanning, *i.e.*, $q_i = \frac{\Omega_i}{\Omega}$, $A = \frac{sN}{\Omega}$ and $B = N$, and Equations (12) and (13) are reduced to Equations (3) and (4), respectively.

C. Discussion

In Equations (12) and (13), A and B are two important factors that control the spreading dynamics of a worm. Meanwhile, A and B are determined by the following parameters: the scanning rate, the vulnerable-host distribution, and the worm-scanning strategy. Thus, Equations (12) and (13) explicitly show how these parameters affect worm spreading. Specifically, when t is small and thus $I(t)$ is small, $[B - I(0)]/[B - I(t)]$ is close to 1, and therefore A dominates the worm propagation speed. It is noted that A is indeed the *infection rate* of a worm that is derived in [4]. As a result, when a worm has a larger infection rate, it can spend much less time to infect the same number of vulnerable hosts at the early stage. Moreover, $\max\{A\} = s \cdot \max_i\{N_i/\Omega_i\}$, *i.e.*, a worm achieves the maximum value of the infection rate when the worm scans only the group containing the largest number of the vulnerable-host density. In this case, the worm uses an extremely non-uniform scanning method. When t and $I(t)$ become larger, B has a greater effect on worm propagation. It is noted that $(\sum_{i=1}^{m} q_i N_i/\Omega_i)^2 \leq N\sum_{i=1}^{m} q_i^2 N_i/\Omega_i^2$ by the Cauchy-Schwarz inequality, which leads to $B \leq N$, where the equality holds if and only if $q_i = \Omega_i/\Omega$, *i.e.*, a worm uses the random scanning. Thus, if a worm uses a more uniform scanning method, B becomes larger and gets close to N, and $[B - I(0)]/[B - I(t)]$ becomes smaller, which leads to smaller t in Equation (12). Therefore, A and B affect worm propagation in very different ways.

It has been observed that if a worm uses a non-uniform scanning method, $B < N$ by the Cauchy-Schwarz inequality. Meanwhile, from Equation (13), it can be seen that $I(t) \leq B$ even when t is very large, assuming $B \geq I(0)$. Thus, for the model described by Equation (13), a worm cannot infect more than B vulnerable hosts. This may not be valid, since



(a) $n = 0$. (b) $n = 0.25$.

(c) $n = 0.5$. (d) $n = 1$.

Fig. 2. Compare the closed-form expression with the extension of the AAWP model.

a worm can infect all N vulnerable hosts under the condition that $q_i > 0$, if $N_i > 0$ for $\forall i$. Therefore, when t is very large, the model may not describe the worm behavior accurately. The reason for this inaccuracy is that in Equation (8), we assume that $q_i u(t)$ is not very large and ignore the higher order terms of the Taylor expansion. Furthermore, based on the above analysis, if a worm uses a more uniform scanning method, B gets closer to N, and our model is more accurate.

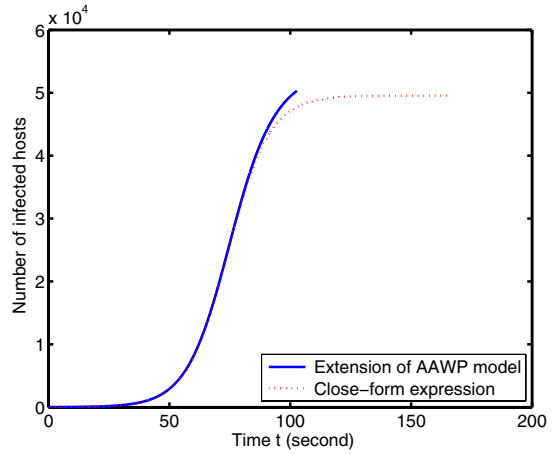
IV. Experimental Results

We evaluate our designed closed-form expression through experiments. In our experiments, we simulate the propagation of a Witty worm that has a vulnerable population $N = 55,909$ and a scanning rate $s = 1,200$ per second [10]. We ignore the effect of disk damages on the Witty worm propagation. We assume that a worm begins spreading from 10 initially infected host (*i.e.*, $I(0) = 10$). In our setting, static worm-scanning strategies exploit the /8 subnets distribution (*i.e.*, $\Omega_1 = \Omega_2 = \cdots = \Omega_{256} = 2^{24}$). Then, we compare worm propagation characterized by the closed-form expression (Equation (13)) with worm spreading described by the extension of the AAWP model (Equation (21) in [4]).

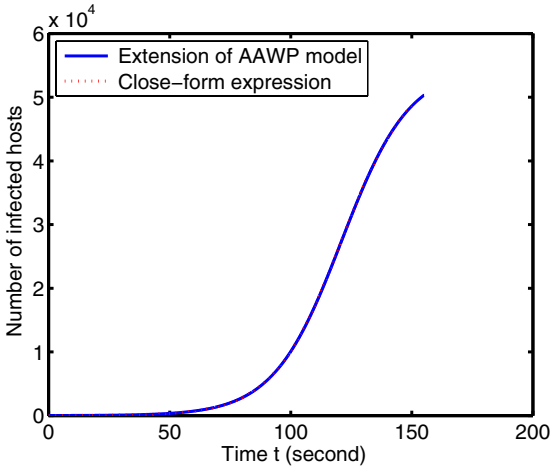
We first consider a group of static worm-scanning strategies where q_i's relate to N_i's explicitly, *i.e.*,

$$q_i = \frac{N_i^n}{\sum_{i=1}^{m} N_i^n} \propto \left(\frac{N_i}{N}\right)^n \tag{14}$$

where $n \geq 0$, representing how strongly q_i depends on N_i/N. If $n = 0$, q_i's are equal and are independent of N_i/N's. In this case, the worm uses random scanning. When n becomes larger, the worm would focus more scans on the groups with a large number of vulnerable hosts, which represents a more non-uniform scanning strategy. Figure 2 compares two worm propagation models when $n = 0$, 0.25, 0.5, and 1. It can be seen that when $n = 0$, *i.e.*, the worm uses random scanning, the curves for both models overlap. When n increases, the closed-form expression describes worm behaviors exactly the same as the extension of the AAWP model before $I(t)$ becomes very large. As we expect, when $I(t)$ is very large, our designed closed-form expression cannot characterize worm dynamics as a result of the effect of the parameter B. Before



(a) OPT-STATIC.



(b) Uniform random sampling of a subset of subnets A.

Fig. 3. Two static scanning strategies from [12].

the infected hosts become saturated, however, the closed-form expression characterizes the average of the number of infected hosts accurately. Moreover, the model is more accurate if the scanning strategy is more uniform (*i.e.*, when n is smaller).

We further apply our model to describe the static strategies proposed in [12]. Specifically, we consider "OPT-STATIC" and "uniform random sampling of a subset of subnets A" strategies that are described in [12]. It can be seen from Figure 3 that our designed closed-form expression can accurately capture the dynamic worm behaviors. Furthermore, we observe that the spreading speeds of these two strategies are very different. While the "OPT-STATIC" strategy takes only 102 seconds to infect 90% vulnerable hosts, the "uniform random sampling of a subset of subnets A" strategy requires 155 seconds. The optimization method proposed in [12], however, cannot characterize this difference in the worm propagation speed.

V. Conclusions

In this paper, we have presented a closed-form expression for modeling the propagation of static worm-scanning strategies. Our model can both accurately characterize the worm propagation speed in the time window of detection and defense and explicitly capture the effects of the vulnerable-host distribution and the worm-scanning method. Therefore, our model can complement with the existing models such as stochastic models [9], [7], deterministic models [16], [2], and optimization methods [12], [3].

As part of our ongoing work, we plan to extend our approach to study the closed-form expressions for modeling the spread of dynamic worm-scanning strategies.

References

[1] Z. Chen, C. Chen, and C. Ji, "Understanding localized-scanning worms," in *Proc. of 26th IEEE International Performance Computing and Communications Conference (IPCCC'07)*, New Orleans, LA, Apr. 2007, pp. 186-193.

[2] Z. Chen, L. Gao, and K. Kwiat, "Modeling the spread of active worms," in *Proc. of INFOCOM'03*, vol. 3, San Francisco, CA, Apr. 2003, pp. 1890-1900.

[3] Z. Chen and C. Ji, "A self-learning worm using importance scanning," in *Proc. ACM/CCS Workshop on Rapid Malcode (WORM'05)*, Fairfax, VA, Nov. 2005, pp. 22-29.

[4] Z. Chen and C. Ji, "Optimal worm-scanning method using vulnerable-host distributions," *International Journal of Security and Networks: Special Issue on Computer and Network Security*, vol. 2, no. 1/2, 2007.

[5] Z. Chen and C. Ji, "Measuring network-aware worm spreading ability," in *Proc. of INFOCOM'07*, Anchorage, AK, May 2007.

[6] Z. Chen, C. Ji, and P. Barford, "Spatial-temporal characteristics of malicious sources," to appear in *Proc. of INFOCOM'08 Mini-Conference*, Phoenix, AZ, April 2008.

[7] D. M. Nicol, "The impact of stochastic variance on worm propagation and detection," in *Proc. ACM/CCS Workshop on Rapid Malcode (WORM'06)*, Fairfax, VA, Nov. 2006.

[8] M. A. Rajab, F. Monrose, and A. Terzis, "On the effectiveness of distributed worm monitoring," in *Proc. of the 14th USENIX Security Symposium (Security'05)*, Baltimore, MD, Aug. 2005, pp. 225-237.

[9] K. Rohloff and T. Basar, "Stochastic behavior of random constant scanning worms," in *Proc. of the 14th ICCCN*, 2005.

[10] C. Shannon and D. Moore, "The spread of the Witty worm," *IEEE Security and Privacy*, vol. 2, no. 4, Jul-Aug 2004, pp. 46-50.

[11] S. Staniford, V. Paxson, and N. Weaver, "How to 0wn the Internet in your spare time," in *Proc. of the 11th USENIX Security Symposium (Security'02)*, San Francisco, CA, Aug. 2002.

[12] M. Vojnovic, V. Gupta, T. Karagiannis, and C. Gkantsidis, "Sampling strategies for epidemic-style information dissemination," to appear in *Proc. of INFOCOM'08*, Phoenix, AZ, April 2008.

[13] J. Wu, S. Vangala, L. Gao, and K. Kwiat, "An effective architecture and algorithm for detecting worms with various scan techniques," in *Proc. 11th Ann. Network and Distributed System Security Symposium (NDSS'04)*, San Diego, CA, Feb. 2004.

[14] C. C. Zou, L. Gao, W. Gong, and D. Towsley, "Monitoring and early warning for Internet worms," in *10th ACM Conference on Computer and Communication Security (CCS'03)*, Washington DC, Oct. 2003.

[15] C. C. Zou, D. Towsley, W. Gong, and S. Cai, "Routing worm: a fast, selective attack worm based on IP address information," in *Proc. 19th ACM/IEEE/SCS Workshop on Principles of Advanced and Distributed Simulation (PADS'05)*, Monterey, CA, June 2005.

[16] C. C. Zou, D. Towsley, and W. Gong, "On the performance of Internet worm scanning strategies," *Elsevier Journal of Performance Evaluation*, vol. 63. no. 7, July 2006, pp. 700-723.

[17] Wikipedia, "Mean field theory," http://en.wikipedia.org/wiki/Mean_field_theory.

[18] Wolfram MathWorld, "Logistic equation," http://mathworld.wolfram.com/LogisticEquation.html.