Heterogeneity in Vulnerable Hosts Slows Down Worm Propagation

Zesheng Chen and Chao Chen
Department of Engineering
Indiana University - Purdue University Fort Wayne, Indiana 46805
Email: {zchen, chen}@engr.ipfw.edu

***Abstract*—Worm attacks continue to be a significant threat to the Internet and have been a main tool used by botnets to recruit bots. Worm propagation models are important for understanding worm dynamics and designing effective and efficient detection and defense systems. The existing models, however, ignore the heterogeneity in vulnerable hosts and assume that the worm-scanning rate is the same for all infected hosts. In this work, we analytically and empirically study the impact of heterogeneity of vulnerable hosts on worm propagation. Specifically, we first apply the Jensen's inequality to show that the heterogeneity in vulnerable hosts indeed hinders the speed of worm propagation. We then conjecture, through the approximation analysis, that if the degree of the heterogeneity in vulnerable hosts is higher, the worm spreads slower. Next, we propose a novel model to predict and characterize worm dynamics among heterogeneous vulnerable hosts. Finally, applying the scale-down simulations and simulating the propagation of a Witty-like worm in the Internet, we verify our analytical results and demonstrate that our proposed model can accurately predict the spread of worms among heterogeneous vulnerable hosts.**

I. Introduction

Worms infect vulnerable hosts and use them to compromise other vulnerable hosts. Such a self-propagation attack has been a significant threat to network security since 2001. Internet worms, such as Code Red, Nimda, Slammer, Witty, and Storm, infected a large number of hosts and caused huge damages. In recent years, worms have also been a main tool used by botnets to recruit a certain number of compromised machines and collect the information of infected hosts. Therefore, it is important and imperative to accurately model the spread of worms in the Internet.

Worm propagation models can help better understand worm dynamic characteristics. More importantly, such models are fundamental for detecting and defending against Internet worms. Mathematical models of worm spreading have been widely studied. For example, differential equations have been used to describe random-scanning worms [13], [19] and to design a worm detection system [18]. A discrete-time model has been proposed with the consideration of host recovery and patch, and has been exploited to monitor, detect, and defend against worms [1]. A stochastic model has been studied to reflect the variation of worm propagation and its impact to worm detection [10]. All existing models, however, assume that vulnerable hosts are homogeneous and as a result, that all infected hosts use the same scanning rate to search for targets. Two related works [17], [6] consider that the scanning rate of infected hosts can vary with time. But these two works also make the assumption that the worm-scanning rate is the same for all infected hosts. Therefore, the impact of heterogeneity in vulnerable hosts on worm propagation has not been studied yet.

Vulnerable hosts in the Internet have been shown to be significantly heterogeneous. The network conditions and the computer performance of end-hosts are *very* different. Specifically, it has been shown that 70% of the end-hosts in a popular BitTorrent system have an upload capacity between 350 Kbps and 1 Mbps, whereas 10% of them have an upload capacity of 10 Mbps or more [5]. Moreover, 64% of the available resources are contributed by only 5% of hosts that have the bandwidth between 55 Mbps and 110 Mbps. A measurement study of the Witty worm also indicates strong heterogeneity in vulnerable hosts [12]. For instance, the bit rates of infected hosts span from less than 56 Kbps to more than 100 Mbps. Hence, when studying worm propagation models, we cannot ignore the effect of the heterogeneity in vulnerable hosts.

The goal of this work is to study the impact of heterogeneity in vulnerable hosts on worm propagation. Specifically, we attempt to answer the following questions:

- Does heterogeneity in vulnerable hosts slow down worm propagation?
- If vulnerable hosts have a higher degree of heterogeneity, would this have a greater impact on worm spreading?
- How can we effectively predict and model worm propagation among heterogeneous vulnerable hosts?

To answer these questions, we analytically and empirically study the worm propagation among both homogeneous and heterogeneous vulnerable hosts. Our analysis is based on the probabilistic model, and the inequality and approximation techniques; whereas the simulation uses the scale-down method and mimics the spread of the Witty-like worm in the Internet. Specifically, we summarize our discoveries and contributions in the following:

- Through both analysis and simulation, we find that statistically the worm has a smaller spreading speed among heterogeneous vulnerable hosts with distinct scanning rates than among homogeneous vulnerable hosts with the same scanning rate. For instance, we demonstrate that a Witty-like worm can be slowed down three times on average in the heterogeneous case than in the homoge-

neous case. Therefore, heterogeneity in vulnerable hosts can potentially slow down worm spreading significantly.
- We show analytically and conjecture that if the degree of heterogeneity in vulnerable hosts is higher, the worm propagates slower. Our simulation results verify the conjecture. This indicates that the current high degree of heterogeneity among vulnerable hosts in the Internet indeed helps defenders to gain some time to respond to worm attacks.
- We then design a novel model to predict the spread of worms among heterogeneous vulnerable hosts. Such a model characterizes the worm propagation delay, *i.e.,* the time difference between the homogeneous case and the heterogeneous case. Simulation results show that our model can accurately predict the dynamics of worm propagation among heterogeneous vulnerable hosts.

The remainder of this paper is structured as follows. Section II discusses the heterogeneity in vulnerable hosts. Section III gives our analysis on worm propagation among heterogeneous vulnerable hosts, whereas Section IV uses simulations to verify our analytical results. Finally, Section V concludes this paper.

II. Heterogeneity in Vulnerable Hosts

Vulnerable hosts in the Internet are heterogeneous. This lies in the fact that end-hosts in the Internet have distinct bandwidth and computer performance. A host may connect to the Internet through a dial-up connection (*e.g.,* 56 Kbps), a digital subscriber line (DSL) (*e.g.,* 4 Kbps $\sim$ 50 Kbps), a local area network (LAN) (*e.g.,* 10 Mbps, 100 Mbps, or 1 Gbps), or a wireless LAN (*e.g.,* 54 Mbps) [7]. Moreover, many worms such as Slammer and Witty are bandwidth limited and send packets as fast as the infected hosts' Internet connection allows [8], [12]. A measurement study on the Witty worm has shown that the infected hosts are heterogeneous [12]. Specifically, while the average transmission speed of an infected host is 3 Mbps, 61% of infected hosts transmit with bit rates between 96 Kbps and 512 Kbps.

For an individual infected host, the bandwidth mainly determines how many scans per unit time a bandwidth-limited worm can send to find targets, *i.e.,* the worm-scanning rate. If an infected host has a higher bandwidth, the worm-scanning rate is always higher. In this work, therefore, we use the variation of worm-scanning rates to reflect the heterogeneity in vulnerable hosts.

III. Theoretical Analysis

Since vulnerable hosts have distinct bandwidth and computer performance, worm-scanning rates from infected hosts can be very different. In this paper, we specifically focus on the impact of the variation of scanning rates on worm propagation and make several simplified assumptions. First, we assume that once a host is infected, it remains in the infection state. Such a susceptible $\rightarrow$ infected (SI) model has been widely used in studying worm spreading [13], [19], [3], [14], [10]. Second, we focus on random-scanning worms. Random scanning selects target IPv4 addresses uniformly and has been exploited by many worms such as Code Red [9], Slammer [8], and Witty [12]. The observations found in this paper, however, can be well extended to other scanning methods such as localized scanning [13] and importance scanning [2]. Finally, while the scanning rates of infected hosts can be different from each other, we assume that the scanning rate of an individual host does not vary with time. This is a reasonable assumption for two reasons: (1) As indicated by our analysis, the time period of worm propagation that we are interested in is at the early stage, *i.e.,* before the worm has infected many hosts and congested networks. (2) It has been observed that an infected host always scans for vulnerable hosts at the maximum speed allowed by its network conditions and computing resources [16].

In this section, we first show theoretically that compared with worm propagation among homogeneous vulnerable hosts, worm spreading is slowed down among heterogeneous vulnerable hosts. We then demonstrate and conjecture that if the degree of heterogeneity in vulnerable hosts is higher, worms spread slower. Finally, we provide a novel worm model that characterizes the spread of worms among heterogeneous vulnerable hosts.

A. Comparing Worm Propagation with Homogeneous Vulnerable Hosts and with Heterogeneous Vulnerable Hosts

We use a discrete-time system to analyze the effect of the variation of scanning rates on worm propagation. Specifically, it is assumed that there are totally N vulnerable hosts and currently I infected hosts. Infected host i ($i = 1, 2, \cdots, I$) uses a scanning rate of s_i, *i.e.,* sends s_i scans per unit time. Then, the total number of scans at the next time step is $\sum_{i=1}^{I} s_i$. Therefore, the probability that an uninfected vulnerable host is hit by a worm scan at the next time step is

$$p_h = \frac{N - I}{\Omega} \cdot \sum_{i=1}^{I} s_i, \tag{1}$$

where Ω is the scanning space. Thus, the time to recruit a new victim, T, follows the geometric distribution, *i.e.,*

$$\Pr(T = k) = p_h(1 - p_h)^{k-1}, \quad k = 1, 2, 3, \cdots \tag{2}$$

which leads to

$$\mathrm{E}[T|s_1, s_2, \cdots, s_I] = \frac{1}{p_h} = \frac{\Omega}{(N - I)\sum_{i=1}^{I} s_i}. \tag{3}$$

It can be seen that if $\mathrm{E}[T]$ is smaller, the worm spreads faster.

If all infected hosts are homogeneous, $s_i = s$, $\forall i$, *i.e.,* the scanning rate for all infected hosts is a constant. Thus, the expected time to recruit a new victim is

$$\mathrm{E}[T] = \frac{\Omega}{sI(N - I)}. \tag{4}$$

On the other hand, if infected hosts are heterogeneous, the scanning rate can be very different for distinct infected hosts. Because of the nature of random scanning, each instant of worm propagation can infect vulnerable hosts in totally different orders. Hence, we assume that s_i's are independent

and identically-distributed (i.i.d.) random variables with mean s and variance σ^2 ($\sigma^2 \geq 0$). Note that if $\sigma^2 = 0$, vulnerable hosts are homogeneous; otherwise, they are heterogeneous. Therefore, from the law of total expectation, we have

$$\mathrm{E}[T] = \mathrm{E}[\mathrm{E}[T|s_1, s_2, \cdots, s_I]] = \frac{\Omega}{N-I}\mathrm{E}\left[\frac{1}{\sum_{i=1}^{I} s_i}\right]. \quad (5)$$

According to the Jensen's inequality [11], [4], if X is a random variable, f is a strictly convex function (*i.e.,* $f''(x) > 0$), and $\mathrm{E}[X]$ and $\mathrm{E}[f(X)]$ exist, then

$$\mathrm{E}[f(X)] \geq f(\mathrm{E}[X]), \quad (6)$$

where the equality holds if an only if X is a constant.

Here, we apply the Jensen's inequality by setting $f(x) = \frac{1}{x}$. Since $f'(x) = -\frac{1}{x^2}$ and $f''(x) = \frac{2}{x^3} > 0$ when $x > 0$, $\frac{1}{x}$ is a strictly convex function. We then find from Equation (5) that

$$\mathrm{E}[T] \geq \frac{\Omega}{N-I} \cdot \frac{1}{\mathrm{E}[\sum_{i=1}^{I} s_i]} = \frac{\Omega}{sI(N-I)}, \quad (7)$$

where the equality holds if and only if $\sigma^2 = 0$.

Comparing Equation (4) and Inequality (7), we have the following theorem.

Theorem 1: If worm-scanning rates s_i's are i.i.d. random variables with mean s and variance σ^2, then the worm spreads slower when $\sigma^2 > 0$ than when $\sigma^2 = 0$. That is, statistically the worm has a smaller spreading speed among heterogeneous vulnerable hosts with distinct scanning rates than among homogeneous vulnerable hosts with the same scanning rate.

Theorem 1 indicates that the existing worm propagation models ignore the variation of scanning rates and thus overestimate the worm propagation speed. Moreover, Theorem 1 reflects that the heterogeneity in vulnerable hosts indeed hinders worm propagation and can help defenders gain some time to respond to worm attacks.

B. Conjecturing the Impact of the Degree of Heterogeneity in Vulnerable Hosts

Since the heterogeneity in vulnerable hosts slows down worm propagation, a question arises: Would the worm spread slower if the degree of the heterogeneity of vulnerable hosts is higher? That is, when σ^2 increases, would $\mathrm{E}[T]$ be larger? To answer this question, we apply Taylor expansion and approximation techniques. Specifically, we study the Taylor expansion of function $f(x) = \frac{1}{x}$, *i.e.,*

$$f(x) = \frac{1}{a} + f'(a)(x-a) + \frac{1}{2}f''(a)(x-a)^2 + \mathrm{H} \quad (8)$$

$$\approx \frac{1}{a} - \frac{x-a}{a^2} + \frac{(x-a)^2}{a^3}. \quad (9)$$

In the above equation, H contains the higher-order terms and can be ignored. Note that $\mathrm{E}[\sum_{i=1}^{I} s_i] = sI$. Then, setting $x = \sum_{i=1}^{I} s_i$ and $a = sI$ in the above equation, we have

$$\frac{1}{\sum_{i=1}^{I} s_i} \approx \frac{1}{sI} - \frac{\sum_{i=1}^{I} s_i - sI}{s^2I^2} + \frac{(\sum_{i=1}^{I} s_i - sI)^2}{s^3I^3}. \quad (10)$$

Taking the expectation on both sides of the above equation, we obtain

$$\mathrm{E}\left[\frac{1}{\sum_{i=1}^{I} s_i}\right] \approx \frac{1}{sI} + \frac{\mathrm{E}[(\sum_{i=1}^{I} s_i - sI)^2]}{s^3I^3} \quad (11)$$

$$= \frac{1}{sI} + \frac{\mathrm{Var}[\sum_{i=1}^{I} s_i]}{s^3I^3} \quad (12)$$

$$= \frac{1}{sI} + \frac{\sigma^2}{s^3I^2}. \quad (13)$$

Therefore, from Equations (5) and (13), the expected time to recruit a new victim is

$$\mathrm{E}[T] \approx \frac{\Omega}{sI(N-I)} + \frac{\Omega\sigma^2}{s^3I^2(N-I)}. \quad (14)$$

In the above equation, the first term (*i.e.,* $\frac{\Omega}{sI(N-I)}$) is identical to $\mathrm{E}[T]$ for the homogeneous case, and the second term is proportional to σ^2. Based on this approximation result, it is obvious that when σ^2 increases, $\mathrm{E}[T]$ also increases. Hence, we have the following conjecture.

Conjecture 1: When σ^2 is larger, the worm spreads slower. That is, the worm propagates slower among the vulnerable hosts with a higher degree of heterogeneity.

C. Modeling Worm Propagation among Heterogeneous Vulnerable Hosts

We apply a novel approach to characterize the spread of random-scanning worms among heterogeneous vulnerable hosts. Instead of obtaining the propagation speed of worms, we attempt to study how much worm propagation delay, compared with the homogeneous case, is caused by the variation of worm-scanning rates. In this way, once we simulate or model the worm spreading among homogeneous vulnerable hosts, we can predict or model the worm propagation among heterogeneous vulnerable hosts.

We first use two worm-scanning rates as an example to demonstrate our modeling procedure. We assume that among N vulnerable hosts, $p \cdot N$ hosts have a scanning rate of r_1, and $(1-p) \cdot N$ hosts have a scanning rate of r_2, where $0 \leq p \leq 1$ and $r_1 \neq r_2$. That is, a randomly selected infected host has a scanning rate of r_1 with probability p and a scanning rate of r_2 with probability $1-p$. Thus, the average scanning rate is $s = pr_1 + (1-p)r_2$. That is, $p = \frac{r_2 - s}{r_2 - r_1}$. Note that p can be derived, given arbitrary values of r_1, r_2, and s. Moreover, among the I infected hosts, the number of hosts having the scanning rate of r_1 follows the binomial distribution $B(I, p)$. If k infected hosts have a scanning rate of r_1, then $\sum_{i=1}^{I} s_i = kr_1 + (I-k)r_2$. From Equation (5), we then obtain

$$\mathrm{E}\left[\frac{1}{\sum_{i=1}^{I} s_i}\right] = \sum_{k=0}^{I} \binom{I}{k} p^k (1-p)^{I-k} \frac{1}{kr_1 + (I-k)r_2}. \quad (15)$$

Therefore, based on the above equation and Equation (4), we can calculate the time difference to recruit a new victim between the heterogeneous case and the homogeneous case,

i.e.,

$$\Delta \mathrm{E}\left[T_I\right] = \sum_{k=0}^{I} \binom{I}{k} \frac{\Omega p^k (1-p)^{I-k}}{[kr_1 + (I-k)r_2](N-I)} - \frac{\Omega}{sI(N-I)}. \quad (16)$$

According to the feature of the binomial distribution, when I is large, $kr_1+(I-k)r_2$ approaches sI with a high probability, and thus $\Delta \mathrm{E}\left[T_I\right]$ is very small and can be ignored. Therefore, we only need to calculate the time difference when I is not large (*e.g.,* $I \leq 1\%$ of the total number of vulnerable hosts). In other words, the worm propagation difference between the heterogeneous case and the homogeneous case only occurs at the early stage of worm spreading when the number of infected hosts is small. Statistically, once a worm has recruited a sufficient number of infected hosts, the heterogeneity in vulnerable hosts has little impact on the worm propagation. On the other hand, when a worm has just started spreading from one or a small number of infected hosts, the impact of the heterogeneity in vulnerable hosts on worm dynamics can be significant, which will be shown in the next section.

Specifically, if we assume that a worm starts spreading from one infected host and set I_0 as the upper bound for calculating $\Delta \mathrm{E}\left[T_I\right]$ in Equation (16), then

$$D_H = \sum_{i=1}^{I_0} \Delta \mathrm{E}\left[T_i\right] \quad (17)$$

represents how much delay is caused by the variation of scanning rates at the early stage of worm propagation. That is, once we obtain the propagation curve for worms among homogeneous vulnerable hosts, we can then shift the curve with the delay D_H to predict the worm spreading among heterogeneous vulnerable hosts with the same average scanning rate.

Note that such a modeling procedure can be easily extended to the case of multiple worm-scanning rates or the case when worm-scanning rates follow an arbitrary distribution. For example, when a worm has multiple scanning rates (*i.e.,* r_1, r_2, $\cdots$, r_m), an infected host has a scanning rate of r_i with probability p_i, where m is the number of scanning rates and $\sum_{i=1}^{m} p_i = 1$. Let n_i ($n_i \geq 0$) denote the number of infected hosts among I infected hosts that have the scanning rate of r_i, where $\sum_{i=1}^{m} n_i = I$. Then, n_i's have a multinomial distribution with parameters I and p_i's, and $\sum_{i=1}^{I} s_i = \sum_{i=1}^{m} n_i r_i$. Therefore, Equation (15) becomes

$$\mathrm{E}\left[\frac{1}{\sum_{i=1}^{I} s_i}\right] = \sum_{\sum n_i = I} \frac{(I!)\left(\prod_{i=1}^{m} p_i^{n_i}\right)}{\left(\prod_{i=1}^{m} n_i!\right)\left(\sum_{i=1}^{m} n_i r_i\right)}. \quad (18)$$

Moreover, if s_i's are i.i.d. random variables with probability distribution $f_S(s)$. Then,

$$\mathrm{E}\left[\frac{1}{\sum_{i=1}^{I} s_i}\right] = \int \cdots \int \frac{\prod_{i=1}^{I} f_S(s_i)}{\sum_{i=1}^{I} s_i} ds_1 \cdots ds_I. \quad (19)$$

In a similar way, we can obtain $\Delta \mathrm{E}\left[T_I\right]$ and D_H for the worm with multiple scanning rates or an arbitrary distribution of scanning rates, and use them to predict the worm propagation among heterogeneous vulnerable hosts.

IV. Simulation Verification

We verify the analytical results in the previous section by simulating the spread of a worm among vulnerable hosts with both homogeneous and heterogeneous scanning rates. As an initial attempt, we only study random-scanning worms with two scanning rates. That is, we assume that some infected hosts have a scanning rate of scan1, whereas others have a scanning rate of scan2. If scan1 = scan2, it is the homogeneous case; otherwise, it is the heterogeneous case. Both homogeneous and heterogeneous cases have the same average worm-scanning rate. Moreover, the target of each worm scan is created by a random number generator over the scanning space, so that each host is hit by the worm scan with an equal probability. Once an uninfected vulnerable host is hit by a worm scan, we record the infection time, *i.e.,* when this vulnerable host is compromised. Based on this infection time, we can count the number of infected hosts at each time step and thus obtain the worm propagation curve. In our simulations, the worm starts spreading from one infected host (*i.e.,* hitlist = 1), which is randomly selected from the vulnerable hosts.

To obtain the analytical results for worm propagation in the heterogeneous case, we first obtain the simulation results for worm spreading in the homogeneous case, and use Equations (16) and (17) to calculate the delay (*i.e.,* D_H) caused by the variation of scanning rates. We then shift the worm propagation curve from the homogeneous case with the delay D_H to predict worm spreading in the heterogeneous case.

Specifically, in this section we first apply scale-down simulations to obtain the observations of worm propagation in a /16 network. We then simulate the spread of Witty worms in the IPv4 address space.

A. Scale-Down Simulations

A scale-down simulation studies worm propagation in a much smaller scanning space, instead of the IPv4 address space that contains 2^{32} IP addresses [15]. In such a way, the patterns of worm spreading can be obtained in a much shorter time through simulations. We apply the technique of scale-down simulations and simulate the spread of random-scanning worms in a /16 subnet. Specifically, we assume that the scanning space is 2^{16} (*i.e.,* $\Omega = 65536$), the number of vulnerable hosts is 5000 (*i.e.,* $N = 5000$), and the average scanning rate is 10 /second (*i.e.,* $s = 10$ /second).

Figure 1 shows the simulation results of worm propagation with four cases of two scanning rates: (1) scan1 = scan2 = 10; (2) scan1 = 5 and scan2 = 15; (3) scan1 = 1 and scan2 = 19; (4) scan1 = 1 and scan2 = 91. The curves in the figure are averages over 10000 runs. It can be seen that compared with the worm in the homogeneous case (*i.e.,* case (1)), worms spread slower in the heterogeneous cases (*i.e.,* cases (2)-(4)), which verifies

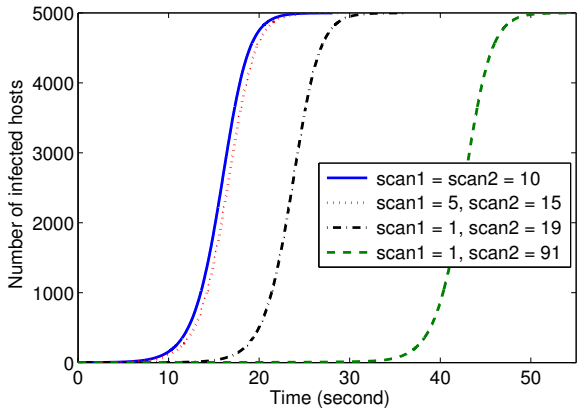


Fig. 1. Impact of scanning-rate variation on worm propagation in scale-down simulations ($\Omega = 65536$, $N = 5000$, $s = 10$ /second, and hitlist = 1).

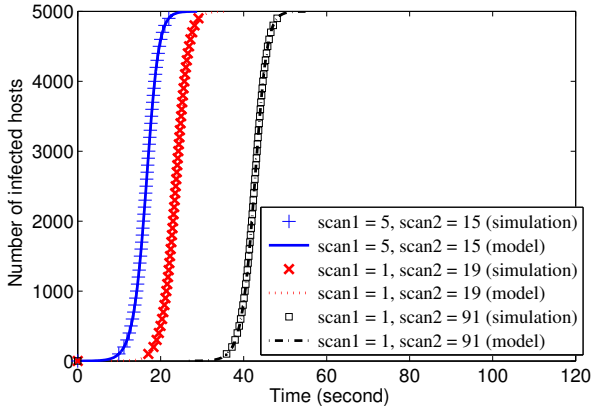


Fig. 2. Comparisons of worm propagation from scale-down simulations and from the model ($\Omega = 65536$, $N = 5000$, $s = 10$ /second, hitlist = 1, and $I_0 = 50$).

Theorem 1. Moreover, if the degree of the heterogeneity in vulnerable hosts is higher, the worm spreads slower, which confirms Conjecture 1. Specifically, the worm takes on average 28.1 seconds to infect all vulnerable hosts in case (1), whereas the worm uses 28.8, 36.0, and 55.0 seconds in cases (2), (3), and (4), respectively. Moreover, it can be seen from the figure that after the worm has infected a certain number of hosts (*e.g.,* 10% of vulnerable hosts), the propagation curves for all four cases are identical, which verifies our observations from Equation (16).

Figure 2 compares the simulation results to our analytical results for the heterogeneous cases. In Equation (17), we set 50 as the upper bound (*i.e.,* $I_0 = 50$) to calculate the delay (*i.e.,* D_H). Specifically, we find that $D_H = 0.7$, 8.1, and 26.8 seconds for cases (2)-(4). From the figure, it can be seen that the curves of analytical results and simulation results overlap, indicating that our prediction is accurate.

B. Witty-Worm Propagation Simulations

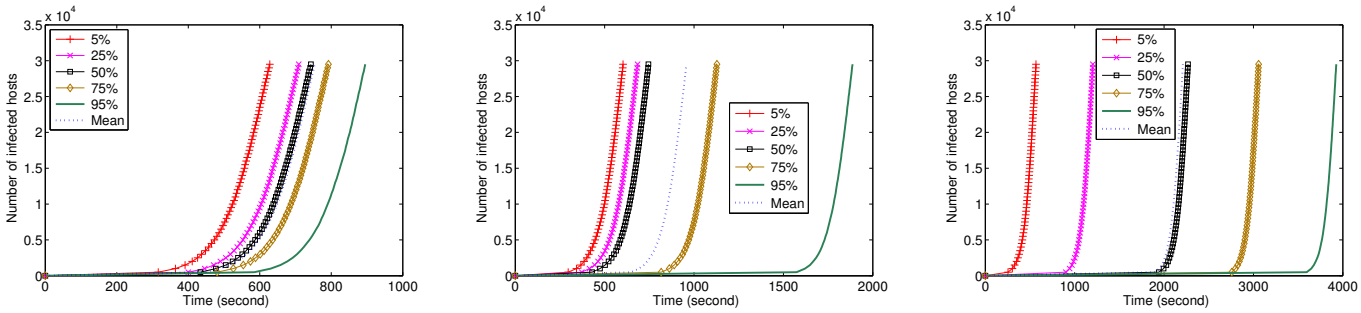
Next, we simulate the spread of a worm in the IPv4 address space, using the parameters from the Witty worm. Specifically, the Witty worm scans the entire IPv4 address space (*i.e.,* $\Omega = 2^{32}$), targets 55909 vulnerable hosts (*i.e.,* $N = 55909$), and uses an average scanning rate of 1200 /seconds (*i.e.,* $s = 1200$ /seconds) [12]. We consider three cases of two worm-scanning rates: (1) scan1 = scan2 = 1200; (2) scan1 = 200 and scan2 = 2200; (3) scan1 = 100 and scan2 = 10000. For case (3), two scanning rates differ 100 times, which is motivated from the observation that the bandwidth capacity of end-hosts can have 100 times difference [5]. For each scenario, we simulate 100 runs with different seeds. Since the major difference among three cases occurs in the time period before the worm infect a significant portion of vulnerable hosts, our simulator stops running when the worm has compromised 30000 hosts.

Figure 3 shows the spread of the Witty worm with three different combinations of two scanning rates. In each sub-figure, the "5%" curve indicates that a worm spreads no faster than this curve in 5 out of 100 simulation runs. The similar definition is applied to the "25%", "50%", "75%", and "95%" curves. Moreover, the "mean" curve is the average over 100 runs. It can be seen that the worm propagates faster in the homogeneous case than in the heterogeneous cases. Furthermore, when the degree of heterogeneity in vulnerable hosts increases, the worm spreads slower, and the variation of worm propagation is larger. These observations are similar to those in the scale-down simulations and verify our analysis. Specifically, comparing cases (1) and (3), we find that the worm uses on average 756.0 seconds to infect 30000 hosts in the homogeneous case, whereas the worm needs 2212.9 seconds to compromise the same number of hosts in the heterogeneous case. This means that the worm is slowed down about 3 times due to the variation of scanning rates and indicates that the heterogeneity in vulnerable hosts can potentially impact worm spreading significantly.

We then further evaluate the performance of our prediction to worm propagation among heterogeneous vulnerable hosts in Figure 4. In our predication, we use only 10 as the upper bound in Equation (17), *i.e.,* $I_0 = 10$. In this figure, the curves of simulations are the averages over 100 runs, whereas the curves of the model are based on Equations (16) and (17). It can also be seen that the curves of simulation and analytical results are very close, indicating that our model well characterizes the dynamics of worm propagation among heterogeneous vulnerable hosts.

V. Conclusions

In this work, we have shown that heterogeneity in vulnerable hosts slows down worm propagation through both analysis and simulation. Moreover, a higher degree of heterogeneity in vulnerable hosts leads to slower propagation of worms. We have also designed a new model to characterize worm spreading among heterogeneous vulnerable hosts. Our model focuses on the worm propagation time difference between the heterogeneous case and the homogeneous case, and is shown empirically to have a good performance to predict worm dynamics. To the best of our knowledge, this is the



(a) Case 1: scan1 = scan2 = 1200 (b) Case 2: scan1 = 200 and scan2 = 2200 (c) Case 3: scan1 = 100 and scan2 = 10000

Fig. 3. Impact of scanning-rate variation on witty-worm propagation ($\Omega = 2^{32}$, $N = 55909$, $s = 1200$ /second, and hitlist = 1).

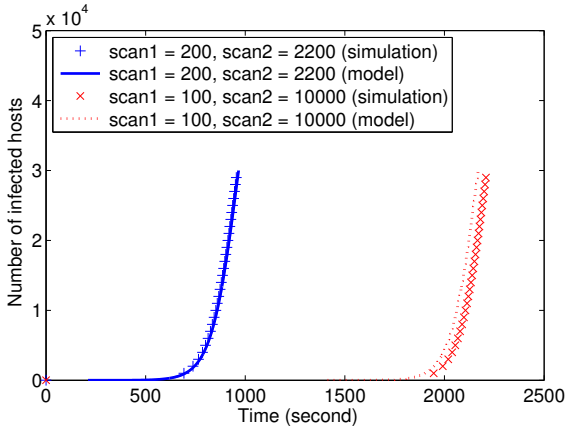


Fig. 4. Comparisons of witty-worm propagation from simulations and from the model ($\Omega = 2^{32}$, $N = 55909$, $s = 1200$ /second, hitlist = 1, and $I_0 = 10$).

first attempt in understanding the impact of the heterogeneity of vulnerable hosts on worm propagation quantitatively.

As our on-going work, we plan to extend the study to other scanning methods such as importance scanning.

REFERENCES

[1] 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.

[2] 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.

[3] Z. Chen and C. Ji, "An information-theoretic view of network-aware malware attacks," *IEEE Transactions on Information Forensics and Security*, vol. 4, no. 3, Sept. 2009, pp. 530-541.

[4] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.

[5] T. Isdal, M. Piatek, A. Krishnamurthy, and T. Anderson, "Leveraging BitTorrent for end host measurements," in *Proc. of the 8th Passive and Active Measurement Conference (PAM '07)*, Louvain-la-neuve, Belgium, Apr. 2007.

[6] E. Kirmani and C. S. Hood, "Analysis of a scanning model of worm propagation," *Journal in Computer Virology*, vol. 6, no. 1, 2010, pp. 31-42.

[7] J. F. Kurose and K. W. Ross, *Computer Networking: A Top-Down Approach*, 4th Editoin, Pearson Education, Inc., 2008.

[8] D. Moore, V. Paxson, S. Savage, C. Shannon, S. Staniford, and N. Weaver, "Inside the Slammer worm," *IEEE Security and Privacy*, vol. 1, no. 4, July 2003, pp. 33-39.

[9] D. Moore, C. Shannon, and J. Brown, "Code-red: a case study on the spread and victims of an Internet worm," in *ACM SIGCOMM/USENIX Internet Measurement Workshop*, Marseille, France, Nov. 2002.

[10] 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.

[11] S. M. Ross, *Stochastic Processes*, Second Edition. John Wiley & Sons, Inc., 1996.

[12] 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.

[13] 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, pp. 149-167.

[14] M. Vojnovic, V. Gupta, T. Karagiannis, and C. Gkantsidis, "Sampling strategies for epidemic-style information dissemination," *IEEE/ACM Transactions on Networking*, vol. 18, no. 4, Aug. 2010, pp. 1013-1025.

[15] N. Weaver, I. Hamadeh, G. Kesidis, and V. Paxson, "Preliminary results using scale-down to explore worm dynamics," in *Proc. of the 2nd ACM Workshop on Rapid Malcode (WORM'04)*, Fairfax, VA, Oct. 2004.

[16] S. Wei and J. Mirkovic, "Correcting congestion-based error in network telescopes observations of worm dynamics," in *Proc. of the 8th Internet Measurement Conference (IMC'08)*, Vouliagmeni, Greece, Oct. 2008.

[17] C. C. Zou, W. Gong, and D. Towsley, "Code red worm propagation modeling and analysis," in *Proc. of the 9th ACM Conference on Computer and Communication Security (CCS'02)*, Washington DC, Nov. 2002, pp. 138-147.

[18] C. C. Zou, W. Gong, D. Towsley, and L. Gao, "The monitoring and early detection of Internet worms," *IEEE/ACM Transactions on Networking*, vol. 13, no. 5, Oct. 2005, pp. 961-974.

[19] 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.