Optimal worm-scanning method using vulnerable-host distributions

Zesheng Chen and Chuanyi Ji
School of Electrical & Computer Engineering
Georgia Institute of Technology, Atlanta, Georgia 30332
Email: {zchen, jic}@ece.gatech.edu

Abstract: Most Internet worms use random scanning. The distribution of vulnerable hosts on the Internet, however, is highly non-uniform over the IP-address space. This implies that random scanning wastes many scans on invulnerable addresses, and more virulent scanning schemes may take advantage of the non-uniformity of a vulnerable-host distribution. Questions then arise as to how attackers may exploit such information and how virulent the resulting worm may be. These issues provide "worst-case scenarios" for defenders and "best-case scenarios" for attackers when the vulnerable-host distribution is available. This work develops such a scenario, called *importance scanning*, which results from importance sampling in statistics. Importance scanning scans the IP-address space according to an empirical distribution of vulnerable hosts. An analytical model is developed to relate the infection rate of worms with the importance-scanning strategies. Based on parameters chosen from Witty and Code Red worms, the experimental results show that an importance-scanning worm can spread much faster than either a random-scanning worm or a routing worm. In addition, a game-theoretical approach suggests that the best strategy for defenders is to scatter applications uniformly in the entire IP-address space.

Keywords: security; worm propagation; modeling; game theory; importance scanning.

1 Introduction

As the number of computers and communication networks increases, Internet worms have become increasingly prevalent [8, 9, 12]. Using malicious, self-propagating codes, worms spread rapidly by infecting computer systems and disseminating themselves in an automated fashion using the infected nodes.

Most worms employ random scanning to select target IP addresses. Since the density of vulnerable hosts is low, a random scan hits a vulnerable machine with a small probability. For example, the Code Red worm infected a vulnerable population of 360,000 machines among 2^{32} IP addresses [19]. Thus, the probability that a random scan will hit a vulnerable target is only $\frac{360{,}000}{2^{32}} = 8.38 \times 10^{-5}$. Therefore, random scanning wastes many scans on invulnerable addresses.

Future worms, however, are likely to employ more effective scanning strategies to identify their targets. Hence, it is important that advanced scanning strategies that can potentially be used to access worst-case scenarios be studied. This work proposes such an optimal scanning method referred to as *importance scanning*. Importance scanning is inspired by importance sampling in statistics [17, 6, 13]. The basic idea of importance sampling is to make rare events occur more frequently and thus reduce the number of samples needed for accurately estimating the corresponding probability. Rare events for worm scanning correspond to hitting a target in a large population. Thus, importance scanning allows attackers to focus on the most relevant parts of an address space so that the probability of hitting vulnerable hosts increases.

Importance scanning relies on a certain statistic of an underlying vulnerable-host distribution. An attacker can potentially obtain such information by querying a database of parties to the vulnerable protocol, stealthy scanning the (partial) target address space, and/or searching the records of old worms [14].

In view of the amount of information an attacker can obtain, random, flash [15], and routing [20] worms can be regarded as special cases of importance-

scanning worms. In particular, a random worm has no information about the vulnerable-host distribution and thus regards the distribution as uniform in the IPv4 space. A flash worm acquires all knowledge, and the target distribution is uniform only in the vulnerable-population space. A routing worm has the knowledge from BGP routing tables about the space of existing hosts, and the corresponding distribution can be considered as uniform in the routing space.

In this work, we assume that a probability distribution of vulnerable hosts is available/obtainable. We then intend to answer the following questions:

- How can an attacker design a fast importance-scanning worm by taking advantage of the knowledge of the vulnerable-host distribution?
- How can we quantitatively analyze the relationship between the speed that worms can achieve and the knowledge that attackers can obtain?
- How can a defender counteract such importance-scanning worms?

To answer these questions, we focus on two quantities: the infection rate that characterizes how fast worms can spread at an early stage and the scanning strategy that is used to locate vulnerable hosts. We first derive relationships between the infection rate and scanning strategies. We then model the spread of importance-scanning worms using the analytical active worm propagation (AAWP) model [1]. We derive the optimal scanning strategy that maximizes the infection rate. That is, the optimal strategy corresponds to the best-case scenario for attackers and the worst-case scenario for defenders. As the optimal strategy is difficult to achieve in reality, we derive a suboptimal scanning strategy as an approximation. To assess the virulence, we compare importance scanning with random and routable scanning. We take the empirical distributions of Witty-worm victims and Web servers as examples of the vulnerable-host distribution. We show that an importance-scanning worm based on parameters chosen from real measurements can spread nearly twice as fast as a routing worm before the victim population becomes saturated.

Moreover, we demonstrate, from the viewpoint of game theory, that a defense mechanism against importance-scanning worms requires the uniform distribution of an application. Under this defense strategy, the best strategy of importance scanning is equivalent to the random-scanning strategy.

Our designed importance scanning is inspired by importance sampling [17, 6, 13]. Our work, however, is different from [17] in that [17] is on estimating the density of Web servers, and we focus on optimal scanning worms that use an uneven vulnerable-host distribution. Hence, while [17] studies a static quantity as the density of Web servers, we consider a dynamic process as the worm propagation. Moreover, [17] uses the variance of an estimator as the performance indicator, and we employ the worm propagation speed, such as the infection rate, as the objective function.

The remainder of this paper is structured as follows. Section 2 provides the background on worm-scanning methods, vulnerable-host distributions, and a random worm propagation model. Section 3 describes the problem. Section 4 characterizes the importance-scanning strategy through the theoretical analysis. Section 5 shows the propagation speed of importance-scanning worms empirically. Section 6 further discusses the defense strategy and Section 7 concludes the paper.

2 Preliminaries

2.1 Scanning Methods

A worm spreads by employing distinct scanning mechanisms such as random, localized, and topological scanning [14]. *Random scanning*, used by such infamous worms as Code Red and Slammer, selects target IP addresses at random. *Localized scanning*, used by the Code Red II and Nimda worms, preferentially scans for hosts in the "local" address space. *Topological scanning*, used by the Morris worm, relies on the "address" information contained in the victim machines to locate new targets.

Some advanced scanning methods have been developed in the research community. For example, Staniford et al. presented the "hitlist" idea [14] to speed up the spread of worms at the initial stage. This list consists of potentially vulnerable machines that are gathered beforehand and targeted first when the worm is released. An extreme case for the hitlist-scanning worms is a *flash worm*, which gathers all vulnerable machines into the list. The flash worms are considered the fastest possible worms [15], for every scan can hit a vulnerable host. Another method that improves the spread of worms is to reduce the scanning space. Attackers can potentially achieve this by using the information provided by BGP routing tables. This type of worm is called a "routable-scanning worm" [16] or

a "routing worm" [20]. Zou et al. designed two types of routing worms [20]. One type, based on Class-A (x.0.0.0/8) address allocations, is thus called "Class-A routing worms." Such worms can reduce the scanning space to 45.3% of the entire IPv4 address space. The other type, based on BGP routing tables, is thus called "BGP routing worms." Such worms can reduce the scanning space to only about 28.6% of the entire IPv4 address space. Another strategy that a worm can potentially employ is *DNS random scanning* [5], in which a worm uses the DNS infrastructure to locate likely targets by guessing DNS names instead of IP addresses. Such a worm in the IPv6 Internet is shown to exhibit a propagation speed comparable to that of an IPv4 random-scanning worm.

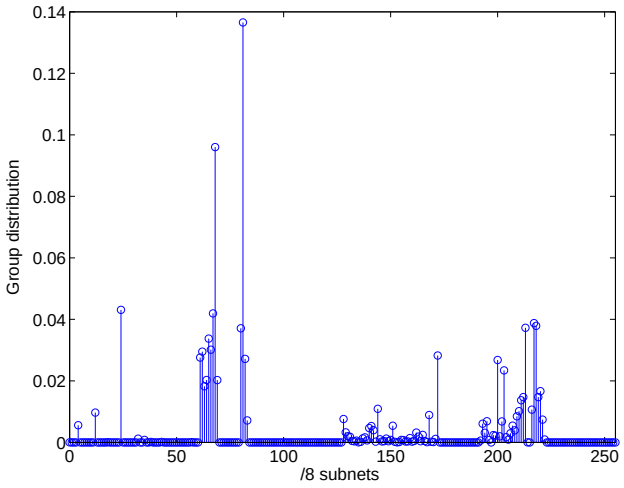
Most of these advanced worms can propagate far faster than a traditional random-scanning worm. When these advanced worms are studied, however, the vulnerable hosts are assumed to be uniformly distributed in either the entire IPv4 address space or the scanning space. Hence, the information on a vulnerable-host distribution is not exploited by the worms.

2.2 Vulnerable-Host Distributions

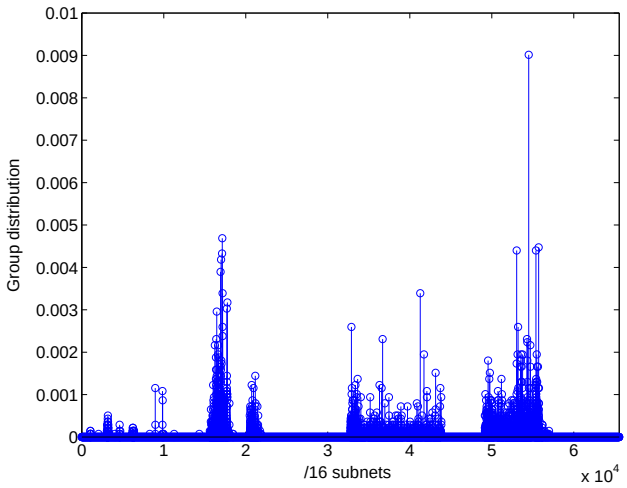
The distribution of vulnerable hosts in the Internet is not uniform. This is evident as no hosts can exist in reserved or multicast IPv4 address ranges assigned by the Internet Assigned Number Authority (IANA) [23, 18]. More importantly, the vulnerable-host distribution may be highly non-uniform over the registered IPv4 address space as indicated by our two collected traces. One trace is a traffic log of the Witty worm obtained from CAIDA [24]. The Witty worm attacks ISS firewall products and carries a destructive payload [12]. CAIDA used a *Network Telescope* to record packets from victims of the Witty worm. Since the network telescope approximately contains the addresses of a Class-A subnet, the collected trace can accurately reflect the distribution of hosts that are vulnerable to the Witty worm [12]. The collected victim addresses are then used to estimate the probability distribution of Witty-worm victims in group i, where

$$p_g(i) = \frac{\text{number of addresses in group } i}{\text{total number of collected addresses}}. \quad (1)$$

The /8 subnet empirical distribution of Witty-worm victims is shown in Figure 1(a). The other trace is on Web servers. We collected 13,866 IP addresses



(a) /8 subnet distribution of Witty-worm victims.



(b) /16 subnet distribution of Web servers.

Figure 1: Uneven distributions of Witty-worm victims and Web servers.

of Web servers provided by the random Uniform Resource Locator (URL) generator from UROULETTE (http://www.uroulette.com/) on January 24, 2005. The empirical distribution of Web servers can be computed via Equation (1), and Figure 1(b) shows the /16 subnet distribution.

It is observed that the distributions of both Witty-worm victims and Web servers are far from uniform in the routable address space. A statistical analysis of network telescope observations also shows that the victims of Code Red and Slammer worms have a highly non-uniform geographical distribution [8, 9]. Moreover, DShield [22] data indicate that the distributions of vulnerable hosts among prefixes follow a *power law* [11].

2.3 Random Worm Propagation Model

We now review a worm propagation model as preparation for relating the rate of worm spread with the distribution of vulnerable hosts. A simple model, known as the *susceptible* → *infected* (SI) model, has been used to model the spread of random-scanning worms in various earlier works [14, 20, 5]. The model assumes that each host has only two states: susceptible or infected. Once infected, a host remains infected.

As importance scanning (sampling) is usually performed in discrete time [17], we adopt a discrete-time SI model. In particular, we use the analytical active worm propagation (AAWP) model, developed by Chen et al. in [1]. In the AAWP model, the spread of random-scanning worms is characterized as follows:

$$I_{t+1} = I_t + (N - I_t)[1 - (1 - \frac{1}{\Omega})^{sI_t}], \qquad (2)$$

where I_t is the number of infected hosts at time t ($t \geq 0$); N is the number of vulnerable hosts; s is the scanning rate of the worm; and Ω is the scanning space. At $t = 0$, I_0 represents the number of hosts on the hitlist.

When a worm begins to spread, $I_t << N$ and $sI_t << \Omega$. The AAWP model can be approximated by

$$I_{t+1} \approx I_t + N \cdot \frac{sI_t}{\Omega} = (1 + \alpha)I_t, \qquad (3)$$

where $\alpha = \frac{sN}{\Omega}$ is the *infection rate* [20]. The infection rate represents the average number of vulnerable hosts that can be infected per unit time by one infected host during the early stage of worm propagation. Based on Equation (3), $I_t \approx (1 + \alpha)^t I_0$, i.e., the number of infected hosts increases exponentially. Therefore, to speed up the spread of worms at the early stage, attackers should design effective scanning methods to increase the infection rate. For instance, a traditional random worm scans the entire IPv4 address space, and thus $\Omega = 2^{32}$. The infection rate of this worm is $\alpha_0 = \frac{sN}{2^{32}}$. In contrast, a BGP and a Class-A routing worm can achieve faster infection rates with the same scanning rate and the same number of targets [20]: $\alpha_1 = \frac{sN}{0.286 \times 2^{32}} = 3.5\alpha_0$ and $\alpha_2 = \frac{sN}{0.453 \times 2^{32}} = 2.2\alpha_0$.

3 Problem Description

We now describe the problems studied in this paper. Let s be the scanning rate or the number of scans that an infected host sends per unit time. Define A_n ($1 \leq n \leq s$) as an IPv4 address probed by the nth scan from an infected host at the early stage of worm propagation. Thus, A_n is a random variable, and $A_n \in \{1, 2, \cdots, 2^{32}\}$. Let $I(A_n)$ denote the vulnerability of address A_n,

$$I(A_n) = \begin{cases} 1, & \text{if address } A_n \text{ is vulnerable to a worm;} \\ 0, & \text{otherwise.} \end{cases}$$

Thus, $\sum_{A_n} I(A_n) = N$. Let $p(A_n)$ denote the actual vulnerable-host distribution, i.e., the probability that $I(A_n) = 1$.

$$p(A_n) = \frac{I(A_n)}{N} = \begin{cases} \frac{1}{N}, & \text{if } I(A_n) = 1; \\ 0, & \text{if } I(A_n) = 0. \end{cases}$$

It is noted that $\sum_{A_n} p(A_n) = 1$.

Let $p^*(A_n)$ denote the probability that the worm scans address A_n. $p^*(A_n)$ can be a uniform distribution as in random-scanning worms or a non-uniform biasing distribution as in flash worms. $p^*(A_n)$ is chosen by an attacker. The choice of the scanning distribution $p^*(A_n)$ is essential to the effectiveness of importance scanning. As we shall see, $p^*(A_n)$ depends on the actual probability distribution $p(A_n)$.

In this paper, we intend to answer the following questions:

- Given complete knowledge about $p(A_n)$, what is the optimal choice of $p^*(A_n)$ that maximizes infection rate α?
- Given partial knowledge about $p(A_n)$, what is the optimal choice of $p^*(A_n)$ that maximizes α?
- What are the spread dynamics of importance-scanning worms using the optimal or the practical choice of $p^*(A_n)$?
- How much faster can an importance-scanning worm spread than a random or a routing worm?
- How can we defend against such importance-scanning worms by customizing $p(A_n)$?

Table 1 shows the notations used in this paper.

4 Importance Scanning

We begin by answering the first three of these five questions in this section. This suffices to deriving the infection rate of importance-scanning worms and modeling the spread of importance-scanning worms.

Table 1: Notations used in this paper.

Notation	Explanation
s	Scanning rate: Number of scans that an infected host sends per unit time
N	Total number of vulnerable hosts
Ω	Scanning space: address space that a worm scans
$p(A_n)$	Actual vulnerable-host distribution: Probability of address A_n being vulnerable to a worm
$p^*(A_n)$	Scanning distribution: Probability of a worm scan hitting address A_n
R	Number of vulnerable hosts that can be infected per unit time by one infected host during the early stage of worm propagation
α	Infection rate: $\alpha = E[R]$
I_t	Expected number of infected host at time t
m	Number of groups in the Internet
N_i	Number of vulnerable hosts in group i
Ω_i	Size of address space in group i
D_i	Set of addresses in network i
$I_{t,i}$	Expected number of infected hosts in group i at time t
$p_g(i)$	Group distribution: Percentage of vulnerable hosts in group i
$p_g^*(i)$	Group scanning distribution: Probability of a worm scan hitting group i
$p_i(b)$	Interface distribution: Probability of finding a vulnerable host with the interface equal to b, given that the host is in network i
$p_i^*(b)$	Interface scanning distribution: Probability of scanning interface b, given that a scan hits network i
v_i	Vulnerable-host density: $v_i = \frac{p_g(i)}{\Omega_i}$

4.1 Infection Rate

Let R be the number of hosts that can be infected per unit time by one infected host during the early stage of worm propagation. R can be expressed as

$$R = \sum_{n=1}^{s} I(A_n), \tag{4}$$

where we assume that different scans do not hit the same target at the early stage of worm propagation, i.e., if $i \neq j$, then $A_i \neq A_j$. Therefore, the infection rate is given by

$$\begin{aligned} \alpha &= E_*[R] && (5)\\ &= \sum_{n=1}^{s} E_*[I(A_n)] && (6)\\ &= \sum_{n=1}^{s} \sum_{A_n} I(A_n)p^*(A_n) && (7)\\ &= N \sum_{n=1}^{s} \sum_{A_n} p(A_n)p^*(A_n) && (8)\\ &= sN \sum_{A_n} p(A_n)p^*(A_n), && (9) \end{aligned}$$

where $E_*[\cdot]$ denotes the expectation with respect to the scanning distribution $p^*(A_n)$. It is noted that

$$\alpha \leq \sum_{n=1}^{s} \sum_{A_n} p^*(A_n) = s, \tag{10}$$

for any $p^*(A_n)$.

Hence, the infection rate is strongly influenced by the choice of scanning distribution $p^*(A_n)$. A choice of $p^*(A_n)$ determines a scanning strategy, and a good choice, in the view of an attacker, should maximize infection rate α. Two special cases have been observed on "choosing" $p^*(A_n)$. The first case is the random-scanning worms, in which $p^*(A_n) = \frac{1}{2^{32}}$. Thus, $\alpha = \frac{sN}{2^{32}} = \alpha_0$. The second case is the flash worms, in which $p^*(A_n) = p(A_n)$. In this case, $p^*(A_n)$ obtains the optimal scanning strategy $p^*_{opt}(A_n)$, which leads to $\max_{p^*(A_n)}\{\alpha\} = s$, indicating that every scan from the worm would hit a vulnerable host.

One interpretation of $p^*_{opt}(A_n)$ suggests that a good worm scanning strategy should concentrate the scans on the areas that are more likely to find a vulnerable host. The vulnerable-host probability distribution $p(A_n)$, however, cannot be obtained without probing the entire IP address space or gathering a complete database of parties to the vulnerable protocol. Therefore, attackers may not acquire the entire knowledge of $p(A_n)$. However, partial knowledge can be obtained, e.g., by aggregating the subspaces of IP addresses.

4.2 Group Distributions

Such partial information is referred to as *group distributions*, which capture the statistics of groups of addresses rather than individual addresses. The vulnerable-host probability distribution in groups is essentially the marginal of the actual distribution $p(A_n)$. Such groups of addresses can be formed in several ways. For example, IP addresses can be grouped by using the conventional 4-byte description. In [17], this approach is applied to measure the size of the Internet via importance sampling. Here, we extract relevant groups in a more general setting by defining the networks. In particular, we regard a *network* as a group of IP addresses that can be identified by such diverse methods as either the first byte of IP addresses (/8 subnets) or IP prefixes in CIDR.

We assume that the Internet is partitioned into m networks. Let D_i $(i = 1, 2, \cdots, m)$ denote the partition set of addresses in network i, which has Ω_i $(\Omega_i \geq 0)$ addresses. Thus, $\sum_{i=1}^{m} \Omega_i = \Omega = 2^{32}$. We define the group distribution $p_g(i)$ $(i = 1, 2, \cdots, m)$ as the proportion of vulnerable hosts in network i, i.e.,

$$p_g(i) = \frac{N_i}{N} = \sum_{A_n \in D_i} p(A_n), \tag{11}$$

where N_i is the population of vulnerable hosts in network i.

The partition of networks reflects the knowledge that attackers can obtain. For example, in one extreme case of random-scanning worms, $m = 1$ and $\Omega_1 = 2^{32}$. In the other extreme case of flash worms, $m = 2^{32}$ and $\Omega_i = 1$ $(i = 1, 2, \cdots, 2^{32})$. Another choice of partitioning networks is based on the first byte of IP addresses (/8 subnets), where $m = 2^8$ and $\Omega_i = 2^{24}$ $(i = 1, 2, \cdots, 2^8)$. The amount of knowledge collected by the worm with the /8 subnet distribution is only partial, somewhere between that by the random worm and that by the flash worm.

Recall that the goal of importance scanning is to maximize the infection rate. From Equation (9), we have the infection rate

$$\alpha = sN \sum_{i=1}^{m} \sum_{A_n \in D_i} p(A_n) p^*(A_n). \tag{12}$$

Refer to the location of an address A_n, which is in network i, as the *interface* denoted by b $(b = 0, 1, \cdots, \Omega_i - 1)$. Let $p_i(b)$ denote the actual probability of finding a vulnerable host with the interface equal to b, given that the host is in network i, i.e., $p_i(b) = \frac{I(A_n)}{N_i}$. Similarly, define *group scanning distribution* $p_g^*(i)$ as the probability of scanning network i and *interface scanning distribution* $p_i^*(b)$ as the probability of scanning interface b, given that a scan hits network i for the scanning distribution $p^*(A_n)$. We can obtain

$$\begin{aligned} p(A_n) &= p_g(i) \cdot p_i(b) && (13) \\ p^*(A_n) &= p_g^*(i) \cdot p_i^*(b), && (14) \end{aligned}$$

where A_n is in network i with interface b. From Equations (13) and (14), the infection rate becomes

$$\begin{aligned} \alpha &= sN \sum_{i=1}^{m} \sum_{b=0}^{\Omega_i - 1} p_g(i) p_i(b) p_g^*(i) p_i^*(b) && (15) \\ &= sN \sum_{i=1}^{m} \left[p_g(i) p_g^*(i) \sum_{b=0}^{\Omega_i - 1} p_i(b) p_i^*(b) \right]. && (16) \end{aligned}$$

We assume that attackers can obtain information only about group distribution $p_g(i)$ and cannot acquire further knowledge about interface distribution $p_i(b)$. Therefore, if a scan hits network i, the Ω_i hosts in this network are targeted by that scan with the same likelihood, i.e., $p_i^*(b) = \frac{1}{\Omega_i}$. Hence, Equation (16) yields

$$\alpha = sN \sum_{i=1}^{m} \frac{p_g(i) p_g^*(i)}{\Omega_i}. \tag{17}$$

Equation (17) provides the relationships among the infection rate, the group distribution, and the group scanning distribution. Let $v_i = \frac{p_g(i)}{\Omega_i}$, referred to as the *vulnerable-host density* in group i, then

$$\begin{aligned} \alpha &= sN \sum_{i=1}^{m} v_i p_g^*(i) && (18) \\ &\leq sN \sum_{i=1}^{m} \max_k \{v_k\} p_g^*(i) && (19) \\ &= sN \max_k \{v_k\}. && (20) \end{aligned}$$

The equality holds when $p_g^*(j) = 1$, $j = \arg\max_k \{v_k\}$; or 0, otherwise. This means that the optimal importance scanning of a worm is to scan only the network with the largest vulnerable-host density.

4.3 Importance-Scanning Worm Propagation Model

We now model the spread dynamics of importance-scanning worms based on the information of a group distribution.

At time t ($t \geq 0$), let $I_{t,i}$ denote the average number of infected hosts in network i. Thus, the total number of infected hosts $I_t = \sum_{i=1}^{m} I_{t,i}$. The rate at which network i is scanned is $sI_t p_g^*(i)$. As an importance scanning worm employs random scanning *within* each network, on the next time epoch, the number of infected hosts in network i can be derived by the AAWP model, i.e.,

$$I_{t+1,i} = I_{t,i} + (N_i - I_{t,i})[1 - (1 - \frac{1}{\Omega_i})^{sI_t p_g^*(i)}], \quad (21)$$

where $i = 1, 2, \cdots, m$ and $t \geq 0$. $I_{0,i}$ is the number of initially infected hosts in network i. The above equation yields

$$I_{t+1,i} = I_{t,i} + sI_t \frac{(N_i - I_{t,i})p_g^*(i)}{\Omega_i} - O(\frac{1}{{\Omega_i}^2}). \quad (22)$$

Since $\frac{1}{\Omega_i} << 1$, we ignore item $O(\frac{1}{{\Omega_i}^2})$. Summing over $i = 1, 2, \cdots, m$ on both sides, we obtain

$$\begin{aligned} I_{t+1} &= I_t + sI_t \sum_{i=1}^{m} (\frac{N_i - I_{t,i}}{\Omega_i}) p_g^*(i) \\ &\leq I_t + sI_t \sum_{i=1}^{m} \max_k \{\frac{N_k - I_{t,k}}{\Omega_k}\} p_g^*(i) \\ &= [1 + s \cdot \max_k \{\frac{N_k - I_{t,k}}{\Omega_k}\}] I_t. \end{aligned} \quad (23)$$

The equality holds when

$$p_g^*(j) = \begin{cases} 1, & j = \arg\max_k \{\frac{N_k - I_{t,k}}{\Omega_k}\}; \\ 0, & \text{otherwise.} \end{cases} \quad (24)$$

When $t = 0$, $N_i >> I_{t,i}$ and then $\max_k \{\frac{N_k - I_{t,k}}{\Omega_k}\} \approx N \max_k \{v_k\}$, which leads to $\alpha = sN \max_k \{v_k\}$. The above derivation results in an optimal importance-scanning strategy that maximizes the infection rate.

Optimal importance scanning:

1. At each time step t, the worm first finds the network that has the largest value of the left vulnerable-host density, i.e., $j = \arg\max_k \{\frac{N_k - I_{t,k}}{\Omega_k}\}$.
2. Then all infected hosts concentrate on scanning this network. That is, $p_g^*(j) = 1$ and $p_g^*(i) = 0$, $\forall i \neq j$.

This optimal importance scanning strategy, however, is not realistic. First, N may not be known in advance. Second, the network that has the largest value of the left vulnerable-host density changes with time, and therefore, the optimal assignment of $p_g^*(i)$ is time-varying. Even when N were given, it would require that each infected host knows $I_{t,i}$, which leads to numerous information exchanges between infected hosts. However, the essence of optimal importance scanning is that it provides the best scenario of worm scanning using the vulnerable-host distribution, which can be used as the baseline for a suboptimal selection of $p_g^*(i)$.

A simple strategy for suboptimal importance scanning is to assume $p_g^*(i) = \frac{p_g(i)/\Omega_i}{\sum_{j=1}^{m} p_g(j)/\Omega_j}$. That is, the probability that a worm scans network i is proportional to the vulnerable-host density of this network. If $\Omega_1 = \Omega_2 = \cdots = \Omega_m$, then $p_g^*(i) = p_g(i)$. For this scanning strategy, Equation (21) becomes

$$I_{t+1,i} = I_{t,i} + (N_i - I_{t,i})[1 - (1 - \frac{1}{\Omega_i})^{sI_t \frac{p_g(i)/\Omega_i}{\sum_{j=1}^{m} p_g(j)/\Omega_j}}]. \quad (25)$$

Suboptimal importance scanning:

1. Before a worm is released, an attacker first obtains vulnerable-host group distribution $p_g(i)$ and then encodes group scanning distribution $p_g^*(i) = \frac{p_g(i)/\Omega_i}{\sum_{j=1}^{m} p_g(j)/\Omega_j}$ in the worm code.
2. At each time step t, the worm scans network i with probability $p_g^*(i)$.

5 Experiments

In this section, we study the propagation speed of importance-scanning worms based on parameters chosen from the real measurements. We first introduce the experimental set-up. We then show the effect of knowledge and vulnerable-host distributions on the propagation of importance-scanning worms. Finally, we compare importance scanning with random and routable scanning.

5.1 Experimental Set-up

In our experiments, we use the model in Equation (2) to imitate the spread of random-scanning and routing worms. Meanwhile, we employ the model in Equations (21) and (24) to study propagation as a result of the optimal importance-scanning strategy. We also use the model in Equation (25) to simulate the spread of suboptimal importance-scanning worms. To implement the models in Equations (21), (24), and (25), we need to obtain group distribution $p_g(i)$. Here, we use

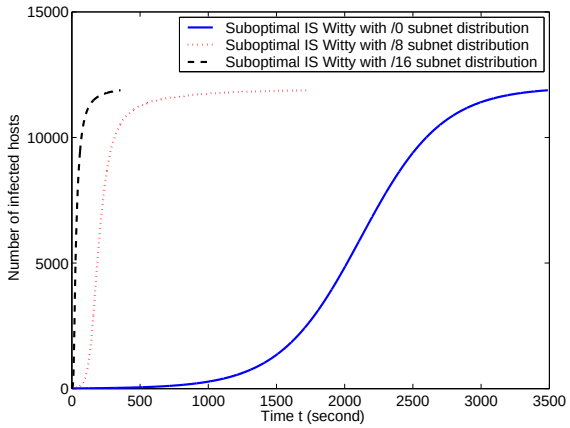


Figure 2: Effect of knowledge.

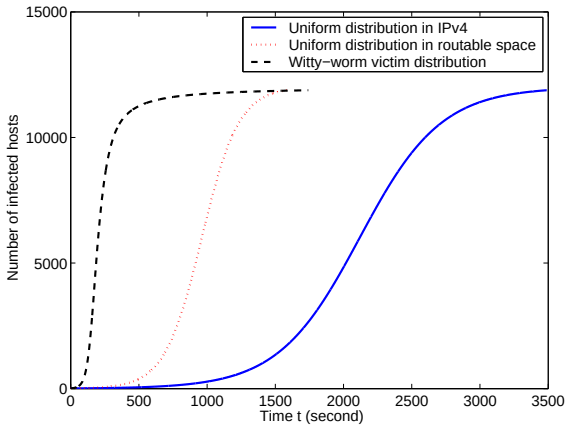


Figure 3: Effect of distributions.

the Witty-worm victim and the Web-server distributions as examples of the vulnerable-host distribution. In other words, we assume that worms attack vulnerable hosts with the same group distribution as that of Witty-worm victims or Web servers. Our collected trace of Web servers does not include all Web servers. However, we assume that the estimated results obtained by Equation (1) are the actual group distribution of Web servers.

The parameters we use in simulated worms are comparable to those in Witty and Code Red worms for evaluating propagation. The Witty worm has a vulnerable population $N = 12,000$ and a scanning rate $s = 1,200$ per second [12, 21]. The Code Red worm has parameters $N = 360,000$ and $s = 358$ per minute [19]. The victims of the Code Red worm is assumed to have the same group distribution as Web servers. We then refer to such an importance-scanning worm as the importance-scanning (IS) Witty or Code Red. Since the experimental results of the Code Red worm are similar to those of the Witty worm, we mainly present the observations from the Witty worm.

5.2 Knowledge Effect

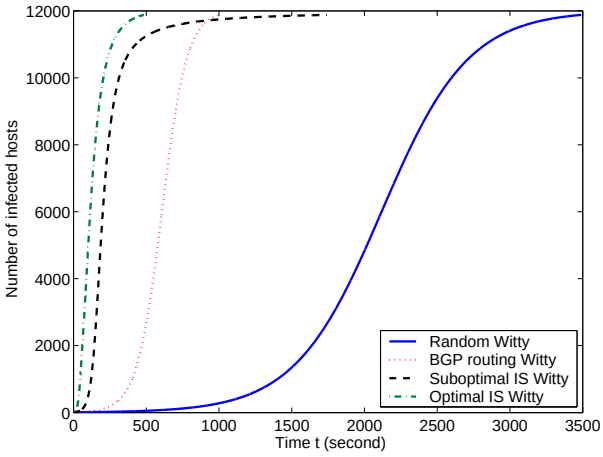
The amount of knowledge about a vulnerable-host distribution affects the rate of spread of importance-scanning worms. Figure 2 shows the propagation comparison among suboptimal importance-scanning Witty worms with different amounts of knowledge about the vulnerable-host distribution, assuming a hitlist of 10 (i.e., $I_0 = 10$). If a worm has the /0 subnet distribution, it knows nothing about the distribution and thus has to use random scanning. We assume that all three Witty worms have the same scanning rate, although a worm that contains more information about the group distribution might slow down for a larger payload. It takes the Witty worm with a /0 subnet distribution 46.3 minutes to infect 90% of vulnerable hosts, whereas the Witty worms with a /8 subnet distribution and a /16 subnet distribution take only 6.6 minutes and 1.6 minutes, respectively. Therefore, more information about the vulnerable-host distribution may help an attacker design a faster worm.

5.3 Vulnerable-Host Distribution Effect

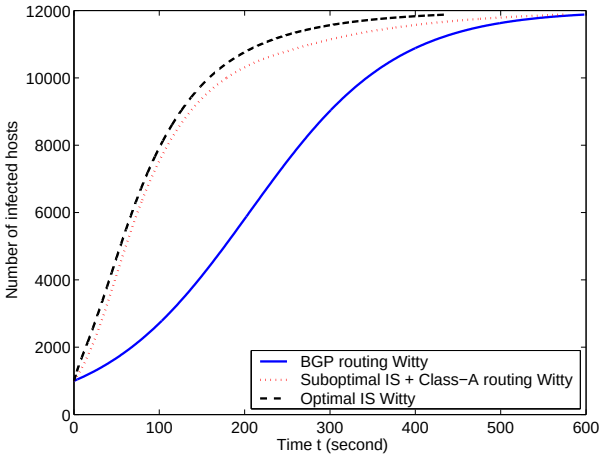
A vulnerable-host distribution also affects the rate of propagation of importance-scanning worms. Figure 3 demonstrates the spread of the suboptimal importance-scanning Witty worms using the /8 subnet distribution, in which vulnerable hosts follow different distributions, assuming a hitlist of 10 (i.e., $I_0 = 10$). A uniform distribution in IPv4 can slow down the worm at least six times than the Witty-worm victim distribution before the victim population becomes saturated. Therefore, the distribution of vulnerable hosts strongly affects the rate of spread of importance-scanning worms.

5.4 Propagation Comparisons

Importance scanning also helps hasten the propagation of a worm. Figure 4(a) shows how propagation as a result of importance-scanning Witty worms compares with that of random and BGP routing Witty worms, assuming a hitlist of 10 (i.e., $I_0 = 10$). The rate of spread of importance-scanning Witty worms increases significantly by using the information on the /8 subnet distribution of vulnerable hosts. The optimal importance-scanning Witty worm can infect 90%



(a) Witty worm with a hitlist of 10



(b) Witty worm with a hitlist of 1,000

Figure 4: Witty worm propagation comparisons.

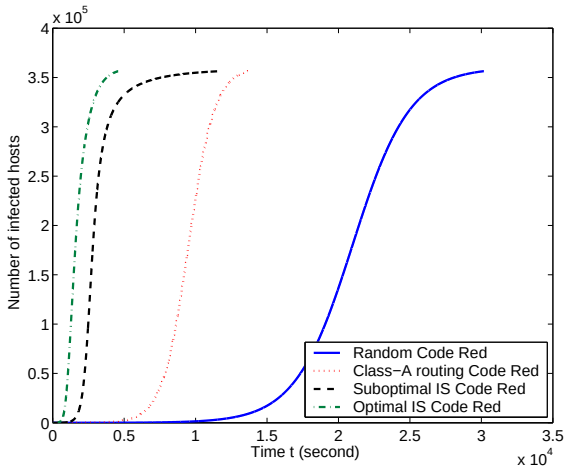


Figure 5: Code Red worm propagation comparison.

vulnerable hosts in as few as 4.2 minutes, whereas the BGP routing Witty worm requires 13.3 minutes. The suboptimal importance-scanning Witty worm spreads more slowly than the optimal worm, but only takes 6.6 minutes to infect the same number of hosts. A BGP routing worm obtains the refined information about the routable space than the worm using the /8 subnet distribution. The BGP routing worm, however, employs random scanning in the BGP routable space. Hence, such a worm, most of time, spreads more slowly than the importance-scanning worm with the /8 subnet distribution, which exploits the underlying uneven distribution of vulnerable hosts.

Once most of the vulnerable hosts are infected, the spread of the suboptimal importance-scanning Witty worm slows down. This is because the suboptimal strategy always uses the same group scanning distribution. As the infected hosts become saturated, a network that initially has more vulnerable hosts actually contains fewer uninfected vulnerable machines. To overcome this problem, suboptimal importance scanning can choose to switch to the routable scanning when only a few uninfected vulnerable hosts are left. Figure 4(b) shows the results for the same experiments, assuming a hitlist of 1,000. Suboptimal importance scanning switches to Class-A routable scanning when 90% vulnerable hosts are infected. Compared with the propagation of a BGP routing worm, importance-scanning worms spread faster before the victim population becomes saturated.

Figure 5 shows the propagation comparison among an optimal importance-scanning Code Red worm, a suboptimal importance-scanning Code Red worm, a Class-A routing Code Red worm, and a random Code Red worm, assuming $I_0 = 10$. The importance-scanning Code Red worms use the /8 subnet distribution. The suboptimal importance-scanning Code Red worm can propagate nearly twice as fast as the Class-A routing Code Red worm before the victim population becomes saturated.

With regard to the storage requirement for /8 subnet group-distribution information, each $p_g(i)$ requires 4 bytes, and each /8 prefix 1 byte. Therefore, the total number of bytes is $5 \times 256 = 1280$. We can reduce this payload by removing the entries with $p_g(i) = 0$, where $i \in \{0, 1, \cdots, 255\}$. Since there are only 97 entries with non-zero $p_g(i)$'s according to the empirical distribution in Figure 1(a), the table can be stored in a $97 \times 5 = 485$ byte payload. Hence, the scanning rate of importance-scanning worms will not decrease much.

6 Game Theory for Attackers and Defenders

Defense against such importance-scanning worms can be modeled by relating it to the interaction between attackers and defenders in game theory. Assume that when an application is introduced to the Internet, defenders can choose how to deploy this application in networks. That is, group distribution $p_g(i)$ can be controlled by defenders, thus leading to a game between attackers and defenders. The attackers attempt to maximize the infection speed (characterized by infection rate α in Equation (17)) by choosing optimal group scanning distribution $p_g^*(i)$, while the defenders endeavor to minimize the worm propagation speed by customizing group distribution $p_g(i)$. Let $V = \{p_g^* : \sum_{i=1}^m p_g^*(i) = 1\}$ stand for the set of group scanning probability vectors p_g^*. Let $U = \{p_g : \sum_{i=1}^m p_g(i) = 1\}$ represent the set of feasible probability assignments for the application distribution. An attacker fears that if a defender knows about the worm-scanning strategy, the defender would then choose a strategy that $\min_{p_g \in U}\{\alpha\}$. Therefore, the objective of an attacker is to choose group scanning distribution $p_g^*(i)$ that maximizes the minimum value, i.e.,

$$\max_{p_g^* \in V} \min_{p_g \in U} sN \sum_{i=1}^{m} \frac{p_g(i)p_g^*(i)}{\Omega_i}. \tag{26}$$

In a similar argument, the objective of a defender is

$$\min_{p_g \in U} \max_{p_g^* \in V} sN \sum_{i=1}^{m} \frac{p_g(i)p_g^*(i)}{\Omega_i}. \tag{27}$$

This is a classical two-person zero-sum game, and the following well-known theorem [10] gives an optimal solution.

Theorem 1. *There exists an optimal solution to the worm-scanning game, where*

$$\alpha_{opt} = \max_{p_g^* \in V} \min_{p_g \in U} sN \sum_{i=1}^{m} \frac{p_g(i)p_g^*(i)}{\Omega_i} \tag{28}$$

$$= \min_{p_g \in U} \max_{p_g^* \in V} sN \sum_{i=1}^{m} \frac{p_g(i)p_g^*(i)}{\Omega_i}, \tag{29}$$

where α_{opt} is the value of the game.

The solution of this *minmax* problem is derived in the following theorem.

Theorem 2. *The value of the worm-scanning game is $\alpha_{opt} = \frac{sN}{2^{32}}$, and the best strategy for a defender is to distribute the application uniformly in the Internet, i.e., $p_g(i) = \frac{\Omega_i}{2^{32}}$, where $i = 1, 2, \cdots, m$.*

PROOF: From Equation (20), we have

$$\max_{p_g^* \in V} \alpha = sN \max_k \{\frac{p_g(k)}{\Omega_k}\}. \tag{30}$$

Set $J = \max_k \{\frac{p_g(k)}{\Omega_k}\}$. The optimal choice of $p_g(i)'s$ requires that J be minimized. Since $\frac{p_g(i)}{\Omega_i} \leq \max_k \{\frac{p_g(k)}{\Omega_k}\} = J$, $p_g(i) \leq J\Omega_i$ for $\forall i$. Thus,

$$1 = \sum_{i=1}^{m} p_g(i) \leq \sum_{i=1}^{m} J\Omega_i = J\Omega, \tag{31}$$

which leads to $J \geq \frac{1}{\Omega}$. The inequality holds when $\frac{p_g(i)}{\Omega_i} = J = \frac{1}{\Omega}$ for $\forall i$. That is, $p_g(i) = \frac{\Omega_i}{\Omega} = \frac{\Omega_i}{2^{32}}$, where $i = 1, 2, \cdots, m$, i.e., the defenders should deploy the application uniformly in the entire IP-address space.

Combining $p_g(i) = \frac{\Omega_i}{2^{32}}$ with Equation (30), the game value is $\alpha_{opt} = \frac{sN}{2^{32}}$. ∎

From Theorem 2, we note that when the defender uses the optimal strategy, the best strategy that the attacker exploits is equivalent to the random-scanning strategy. Meanwhile, Figure 3 demonstrates that the vulnerable-host distribution has a strong effect on worm propagation. Therefore, the design of the future Internet should consider how to distribute an application in security engineering.

In the current Internet, however, the application distributor may not control how to deploy the application in the entire IPv4 address space. Although not applicable for the entire Internet, the best strategy of defenders can still apply for enterprise networks. That is, if an enterprise network attempts to defend against importance-scanning worms, the administrator of this network should distribute the application uniformly in the entire enterprise network from the viewpoint of game theory.

7 Conclusions

In order to effectively defend against Internet worms, we must study potential scanning techniques that attackers may employ. In this paper, we present an optimal worm-scanning method, called *importance scanning*, using the information of a vulnerable-host distribution. This scanning strategy then provide a best-case scenario for attackers when the vulnerable-host distribution is available. Importance scanning can be combined with other scanning methods such as hitlist scanning. Moreover, the division of groups

can be very general, such as Domain Name System (DNS) Top-Level Domains, countries, Autonomous Systems, IP prefixes in Classless Inter-Domain Routing (CIDR), the first byte of IP addresses (/8 subnets), or first two bytes of IP addresses (/16 subnets). For example, when naming distribution information is exploited, importance scanning can also be applied to DNS worms [5], which is worth further investigation. In addition, when IPv4 is updated to IPv6, an importance-scanning worm will not be slowed down very much if vulnerable hosts are still distributed in a clustered fashion. A game-theoretical approach suggests that the best strategy for defenders is to distribute the applications evenly in the entire address space or in each enterprise network.

As part of our ongoing work, we are studying how an intelligent worm can learn about the underlying vulnerable-host distribution if such information is unknown before the worm is released [3]. We plan to investigate how future worms can use advanced learning algorithms to obtain information needed and how we can counteract such smart worms.

ACKNOWLEDGEMENTS

The authors would like to thank Colleen Shannon at CAIDA and Shubho Sen at AT&T labs for kind help. The authors also thank CAIDA for making the Witty dataset available. Support from NSF ECS 0300605 is gratefully acknowledged.

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, "Importance-scanning worm using vulnerable-host distribution," in *Proc. 48th Ann. IEEE Global Telecommunications Conference (GLOBECOM'05)*, St. Louis, MO, Nov. 2005.

[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] D.J. Daley and J. Gani, *Epidemic Modelling: An Introduction*. Cambridge, UK: Cambridge University Press, 1999.

[5] H. Feng, A. Kamra, V. Misra, and A. D. Keromytis, "The effect of DNS delays on worm propagation in an IPv6 Internet," in *Proc. of INFOCOM'05*, vol. 4, Miami, FL, Mar. 2005, pp. 2405-2414.

[6] P. Heidelberger, "Fast simulation of rare events in queueing and reliability models," *ACM Transactions on Modeling and Computer Simulation*, vol.5, no.1, Jan. 1995, pp. 43-85.

[7] J. Ma, G. M. Voelker, and S. Savage, "Self-stopping worms," in *Proc. ACM/CCS Workshop on Rapid Malcode (WORM'05)*, Fairfax, VA, Nov. 2005, pp. 12-21.

[8] 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.

[9] 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.

[10] G. Owen, *Game Theory*. New York: Academic Press, 2001.

[11] 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.

[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] P.J. Smith, M. Shafi, and H. Gao, "Quick simulation: a review of importance sampling techniques in communications systems," *IEEE Jour. Selected Areas Commun.*, vol.15, May 1997, pp.597-613.

[14] 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.

[15] S. Staniford, D. Moore, V. Paxson, and N. Weaver, "The top speed of flash worms," in *Proc. ACM/CCS Workshop on Rapid Malcode (WORM'04)*, Washington DC, Oct. 2004, pp. 33-42.

[16] 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.

[17] S. Xing and B.-P. Paris, "Measuring the size of the Internet via importance sampling," *IEEE journal on selected areas in communications*, vol. 21, no. 6, Aug. 2003, pp. 922-933.

[18] S. Xing and B.-P. Paris, "Mapping the Growth of the Internet," *Proc. of IEEE 2003 International Conference on Computer Communications and Networks*, Dallas, TX, Oct. 2003, pp 199-204.

[19] 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.

[20] 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.

[21] C. C. Zou, D. Towsley, and W. Gong, "On the performance of Internet worm scanning strategies," *to appear in Journal of Performance Evaluation.*

[22] Distributed Intrusion Detection System (DShield) [Online]. Available: http://www.dshield.org/.

[23] Internet Protocol V4 Address Space [Online]. Available: http://www.iana.org/assignments/ipv4-address-space.

[24] The CAIDA Dataset on the Witty Worm - March 19-24, 2004, Colleen Shannon and David Moore, http://www.caida.org/passive/witty/. Support for the Witty Worm Dataset and the UCSD Network Telescope are provided by Cisco Systems, Limelight Networks, the US Department of Homeland Security, the National Science Foundation, and CAIDA, DARPA, Digital Envoy, and CAIDA Members.