Measuring Network-Aware Worm Spreading Ability

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

***Abstract*— This work investigates three aspects: (a) a *network vulnerability* as the non-uniform vulnerable-host distribution, (b) *threats*, i.e., intelligent worms that exploit such a vulnerability, and (c) *defense*, i.e., challenges for fighting the threats. We first study five data sets and observe consistent clustered vulnerable-host distributions. We then present a new metric, referred to as the *non-uniformity factor*, which quantifies the unevenness of a vulnerable-host distribution. This metric is essentially the Renyi information entropy and better characterizes the non-uniformity of a distribution than the Shannon entropy. We then analytically and empirically measure the infection rate and the propagation speed of network-aware worms. We show that a representative network-aware worm can increase the spreading speed by exactly or nearly a non-uniformity factor when compared to a random-scanning worm at the early stage of worm propagation. This implies that when a worm exploits an uneven vulnerable-host distribution as a network-wide vulnerability, the Internet can be infected much more rapidly. Furthermore, we analyze the effectiveness of defense strategies on the spread of network-aware worms. Our results demonstrate that counteracting network-aware worms is a significant challenge for the strategies that include host-based defense and IPv6.**

I. Introduction

Worm scanning has become more and more sophisticated since the initial attacks of Internet worms. Most of the real, especially "old" worms, such as Code Red [12], Slammer [13], and latter Witty [19], exploit naive random scanning that chooses target IP addresses uniformly and does not use any information on network vulnerabilities. Advanced scanning methods, however, have been developed that take the IP address structure into consideration. One example is routable scanning that selects targets only in the routable address space, using the information provided by the BGP routing table [23], [26]. Another example is evasive worms that exploit lightweight sampling to obtain the knowledge of *live* subnets of the address space and spread only in these networks [16].

This work focuses on a class of *network-aware worms*. Such worms exploit the information on the highly uneven distributions of vulnerable hosts. The vulnerable-host distributions have been observed to be bursty and spatially inhomogeneous by Barford et al. [1]. A non-uniform distribution of Witty-worm victims has been reported by Rajab et al. [15]. We have also found that a Web-server distribution is non-uniform in the IP address space [6]. These discoveries suggest that vulnerable hosts and Web servers may be "clustered" (i.e., non-uniform). The clustering/non-uniformity makes the network vulnerable since if one host is compromised in a cluster, the rest may be compromised rather quickly.

In our prior work, we have studied a class of "worst-case" worms, called *importance-scanning* worms, which exploit non-uniform vulnerable-host distributions [6], [5]. Importance scanning is developed from and named after importance sampling in statistics. Importance scanning probes the Internet according to an underlying vulnerable-host distribution. Such a scanning method forces worm scans on the most relevant parts of an address space and supplies the optimal strategy[1]. Importance scanning thus provides a "what-if" scenario: When there are many ways for intelligent worms to exploit such a vulnerability, importance scanning is a worst-case threat-model. Hence, importance scanning can serve as a benchmark for studying real worms.

Are there any real network-aware worms? Code Red II and Nimda worms have used localized scanning [28], [29]. Localized scanning preferentially searches for vulnerable hosts in the "local" address space. The Blaster worm has used sequential scanning in addition to localized scanning [31]. Sequential scanning searches for vulnerable hosts through their closeness in the IP address space. It is not well understood, however, how to characterize the relationships between vulnerable-host distributions and these network-aware worms.

What has been observed is that real network-aware and importance-scanning worms spread much faster than random-scanning worms [15], [6]. This shows the importance of the problem. Does there exist a *generic* characteristic across different vulnerable-host distributions? If so, how do intelligent worms exploit such a vulnerability, and how can we defend against such worms?

Our goal is to investigate such a generic characteristic in vulnerable-host distributions, to quantify its relationship with network-aware worms, and to understand the effectiveness of defense strategies. In particular, we would like to answer the following questions:

- How to quantify the non-uniformity of a vulnerable-host distribution by a simple metric?
- How to measure the spreading ability of network-aware worms quantitatively?
- How to relate vulnerable-host distributions with network-aware worm spreading ability?
- What are the challenges to defense strategies on slowing down the spread of a network-aware worm?

To answer these questions, we first observe, from five measurement sets, common characteristics of non-uniform vulnerable-host distributions. We then derive a new metric as the *non-uniformity factor* to characterize the non-uniformity of a vulnerable-host distribution. A larger non-uniformity factor

[1]Hitlist scanning [21] can be regarded as a special case of importance scanning when the complete information of vulnerable hosts is known.

reflects a more non-uniform distribution of vulnerable hosts. We obtain the non-uniformity factors from the data sets on vulnerable-host distributions and show that all data sets have large non-uniformity factors. Moreover, the non-uniformity factor is a function of the Renyi entropy, a generalized entropy, of order two [17]. We show that the non-uniformity factor better characterizes the unevenness of a distribution than the Shannon entropy. Therefore, in view of information theory, the non-uniformity factor provides a quantitative measure of the unevenness/uncertainty of a vulnerable-host distribution.

Next, we analyze the spreading speed of network-aware worms, especially at an early stage. A worm that spreads faster at the early stage can in general infect most of the vulnerable hosts in a shorter time. The propagation ability of a worm at the early stage is characterized by the *infection rate* [26]. Therefore, we derive the infection rates of network-aware worms. We find that the infection rates of representative network-aware worms can be represented explicitly as a function of the non-uniformity factor. For example, localized scanning can increase the infection rate by nearly a non-uniformity factor, comparing to random scanning. Thus, the spreading speed of localized scanning can approach the capacity of suboptimal importance scanning [6]. These analytical results on the relationships between vulnerable-host distributions and network-aware worm spreading ability are validated by simulation. Furthermore, to show the generality of our approach, we study sequential scanning. We demonstrate that a combination of sequential scanning and random scanning can increase the infection rate significantly.

Finally, we study new challenges to worm defense posed by network-aware worms. Using the non-uniformity factor, we show quantitatively that the host-based defense strategies, such as proactive protection [3] and virus throttling [22], should be deployed at almost all hosts to slow down network-aware worms at the early stage. A partial deployment would nearly invalidate such host-based defense. Moreover, we demonstrate that the infection rate of a network-aware worm in the IPv6 Internet can be comparable to that of the Code Red v2 worm in the IPv4 Internet. Therefore, fighting network-aware worms is a real challenge.

The remainder of this paper is structured as follows. Section II presents our collected data sets. Sections III and IV introduce a new metric called the non-uniformity factor and compare this metric to the Shannon entropy. Sections V and VI characterize the spreading ability of network-aware worms through theoretical analysis and simulations. Section VII further studies the effectiveness of defense strategies on network-aware worms. Section VIII concludes this paper.

II. Measurements and Vulnerable-Host Distribution

How significant is the unevenness of vulnerable-host distributions? To answer this question, we study five data sets.

A. Measurements

DShield (D1): DShield collects intrusion detection system (IDS) logs [30]. Specifically, DShield provides the information of vulnerable hosts by aggregating logs from more than 1,600 IDSes distributed throughout the Internet. We further focus on the following ports that were attacked by worms: 80 (HTTP), 135 (DCE/RPC), 445 (NetBIOS/SMB), 1023 (FTP servers and the remote shell attacked by W32.Sasser.E.Worm), and 6129 (DameWare).

iSinks (P1 and C1): Two unused address space monitors run the *iSink* system [24]. The monitors record the unwanted traffic arriving at the unused address spaces that include a Class A network (referred to as "Provider" or P1) and two Class B networks at the campus of the University of Wisconsin (referred to as "Campus" or C1) [1].

Witty-worm victims (W1): A list of Witty-worm victims is provided by CAIDA [19]. CAIDA used a network telescope with approximate 2^{24} IP addresses to log the traffic of Witty-worm victims that are Internet security systems (ISS) products.

Web-server list (W2): IP addresses of Web servers were collected through UROULETTE (http://www.uroulette.com/). UROULETTE provides a random uniform resource locator (URL) generator to obtain a list of IP addresses of Web servers.

The first three data sets (D1, P1, and C1) were collected over a seven-day period from 10-16 December 2004 and have been studied in [1] to demonstrate the bursty and spatially inhomogeneous distribution of (malicious) source IP addresses. The last two data sets (W1 and W2) have been used in our prior work [6] to show the virulence of importance-scanning worms. The summary of our data sets is given in Table I.

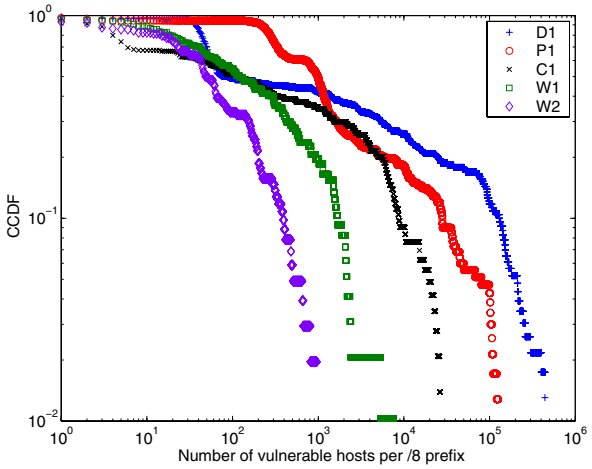
TABLE I
SUMMARY OF THE DATA SETS.

Trace	Description	# of unique source addresses
D1	DShield	7,694,291
P1	Provider	2,355,150
C1	Campus	448,894
W1	Witty-worm victims	55,909
W2	Web servers	13,866

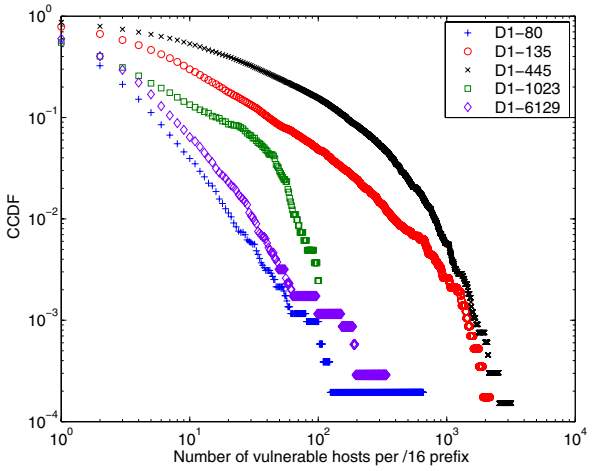
B. Vulnerable-Host Distribution

To obtain vulnerable-host group distributions, we use the classless inter-domain routing (CIDR) notation [11]. The Internet is partitioned into subnets according to the first l bits of IP addresses, i.e., /l prefixes or /l subnets. In this division, there are 2^l subnets, and each subnet contains 2^{32-l} addresses, where $l \in \{0, 1, \cdots, 32\}$. For example, when $l = 8$, the Internet is grouped into Class A subnets (i.e., /8 subnets); when $l = 16$, the Internet is partitioned into Class B subnets (i.e., /16 subnets).

We plot the complementary cumulative distribution functions (CCDF) of our collected data sets in /8 and /16 subnets in Figure 1 in log-log scales. CCDF is defined as the faction of the subnets with the number of hosts greater than a given value. Figure 1(a) shows population distributions in /8 subnets for D1, P1, C1, W1, and W2, whereas Figure 1(b) exhibits host distributions in /16 subnets for D1 with different ports (80, 135, 445, 1023, and 6129). Figure 1 demonstrates a wide range of populations, indicating highly inhomogeneous address structures. Specifically, the relatively straight lines,



(a) Population distributions in /8 subnets.



(b) Population distributions in /16 subnets.

Fig. 1. CCDF of collected data sets.

such as W2 and D1-135, imply that vulnerable hosts follow a power law distribution. Similar observations were given in [1], [15], [14], [12], [13], [6].

Why is the vulnerable-host distribution non-uniform in the IPv4 address space? First, no vulnerable hosts can exist in reserved or multicast address ranges [32]. Second, different subnet administrators make different use of their own IP address space. Third, a subnet intends to have many computers with the same operating systems and applications for easy management [20], [4]. Last, some subnets are more protected than others [1], [15].

How can we quantify the non-uniformity of a vulnerable-host distribution? One way is to use the population distribution such as CCDF plotted in Figure 1. But it is complex to compare the unevenness of two distributions.

III. Non-Uniformity Factor

In this section, we derive a simple metric, called the *non-uniformity factor*, to quantify the non-uniformity of a vulnerable-host distribution.

A. Definition and Property

Let $p_g^{(l)}(i)$ $(i = 1, 2, \cdots, 2^l)$ denote the group distribution of vulnerable hosts in /l subnets. Let $N_i^{(l)}$ be the number of vulnerable hosts in /l subnet i and N be the total number of vulnerable hosts. Then, $p_g^{(l)}(i) = \frac{N_i^{(l)}}{N}$, which is the ratio between the number of vulnerable hosts in group i and the total number of vulnerable hosts. It is noted that $\sum_{i=1}^{2^l} p_g^{(l)}(i) = 1$ and $\sum_{i=1}^{2^l} N_i^{(l)} = N$.

Definition: The *non-uniformity factor* in /l subnets is defined as

$$\beta^{(l)} = 2^l \sum_{i=1}^{2^l} \left(p_g^{(l)}(i)\right)^2. \tag{1}$$

It is noted that

$$\beta^{(l)} \geq \left(\sum_{i=1}^{2^l} p_g^{(l)}(i)\right)^2 = 1. \tag{2}$$

The above inequality is derived by the Cauchy-Schwarz inequality. The equality holds if and only if $p_g^{(l)}(i) = 2^{-l}$, for $i = 1, 2, \cdots, 2^l$. In other words, when the vulnerable-host distribution is uniform, $\beta^{(l)}$ achieves the minimum value 1. On the other hand, since $p_g^{(l)}(i) \geq 0$,

$$\beta^{(l)} \leq 2^l \cdot \left(\sum_{i=1}^{2^l} p_g^{(l)}(i)\right)^2 = 2^l. \tag{3}$$

The equality holds when $p_g^{(l)}(j) = 1$ for some j and $p_g^{(l)}(i) = 0$, $i \neq j$, i.e., all vulnerable hosts concentrate on one subnet. This means that when the vulnerable-host distribution is extremely non-uniform, $\beta^{(l)}$ obtains the maximum value 2^l. Therefore, $\beta^{(l)}$ characterizes the non-uniformity of a vulnerable-host distribution. A larger non-uniformity factor reflects a more non-uniform distribution of vulnerable hosts.

How does $\beta^{(l)}$ vary with l? When $l = 0$, $\beta^{(0)} = 1$. In the other extreme where $l = 32$,

$$p_g^{(32)}(i) = \begin{cases} \frac{1}{N}, & \text{address } i \text{ is vulnerable to the worm;} \\ 0, & \text{otherwise,} \end{cases} \tag{4}$$

which results in $\beta^{(32)} = \frac{2^{32}}{N}$. More importantly, $\beta^{(l)}$ is a non-decreasing function of l, as shown below.

Theorem 1: If $l > r$, $\beta^{(l)} \geq \beta^{(r)}$, where $l, r \in \{0, 1, \cdots, 32\}$.

Proof: Let $k = l - r$. Group i $(i = 1, 2, \cdots, 2^r)$ of /r subnets is partitioned into groups $2^k \cdot (i-1) + 1$, $2^k \cdot (i-1) + 2$, $\cdots$, $2^k \cdot (i-1) + 2^k$ of /l subnets. Thus,

$$p_g^{(r)}(i) = \sum_{j=1}^{2^k} p_g^{(l)}(2^k \cdot (i-1) + j), \qquad i = 1, 2, \cdots, 2^r. \tag{5}$$

Then, $\beta^{(l)}$ is related to $\beta^{(r)}$ by the Cauchy-Schwarz inequality.

$$\begin{aligned}\beta^{(l)} &= 2^r \sum_{i=1}^{2^r} \left\{ \left(\sum_{j=1}^{2^k} 1^2 \right) \left[\sum_{j=1}^{2^k} \left(p_g^{(l)}(2^k \cdot (i-1)+j) \right)^2 \right] \right\} \\ &\geq 2^r \sum_{i=1}^{2^r} \left(\sum_{j=1}^{2^k} p_g^{(l)}(2^k \cdot (i-1)+j) \right)^2 \\ &= \beta^{(r)}. \end{aligned} \quad (6)$$

The equality holds when $p_g^{(l)}(2^k \cdot (i-1)+j) = \frac{p_g^{(r)}(i)}{2^k}$, $j = 1, 2, \cdots, 2^k$, $i = 1, 2, \cdots, 2^r$. That is, in each /r subnet, the vulnerable hosts are uniformly distributed in 2^k groups. ∎

An intuitive explanation of this theorem is as follows. For /l and /$(l+1)$ subnets, group i ($i = 1, 2, \cdots, 2^l$) of /l subnets is partitioned into groups $2i-1$ and $2i$ of /$(l+1)$ subnets. If vulnerable hosts in each group of /l subnets are equally divided into groups of /$(l+1)$ subnets (i.e., $p_g^{(l+1)}(2i-1) = p_g^{(l+1)}(2i) = \frac{1}{2} p_g^{(l)}(i)$, $\forall\ i$), then $\beta^{(l+1)} = \beta^{(l)}$. Otherwise, if the division of vulnerable hosts is uneven for a group (i.e., $p_g^{(l+1)}(2i-1) \neq p_g^{(l+1)}(2i)$, $\exists\ i$), then $\beta^{(l+1)} > \beta^{(l)}$.

B. Estimated Non-Uniformity Factor

Figure 2 shows the non-uniformity factors estimated from our data sets. The non-uniformity factors increase with the prefix length for all data sets. The y-axis is in a *log* scale. Thus, $\beta^{(l)}$ increases *almost exponentially* with a wide range of l. To gain intuition on how large $\beta^{(l)}$ can be, $\beta^{(8)}$ and $\beta^{(16)}$ are summarized for all data sets in Table II. We observe that $\beta^{(8)}$ and $\beta^{(16)}$ have large values, indicating the significant unevenness of collected distributions.

TABLE II

$\beta^{(8)}$ AND $\beta^{(16)}$ OF COLLECTED DISTRIBUTIONS.

$\beta^{(l)}$	D1	P1	C1	W1	W2
$\beta^{(8)}$	7.9	8.4	9.0	12.0	7.8
$\beta^{(16)}$	31.2	43.2	52.2	126.7	50.2
$\beta^{(l)}$	D1-80	D1-135	D1-445	D1-1023	D1-6129
$\beta^{(8)}$	7.9	15.4	10.5	48.2	9.1
$\beta^{(16)}$	153.3	186.6	71.7	416.3	128.9

IV. Entropy and Non-Uniformity Factor

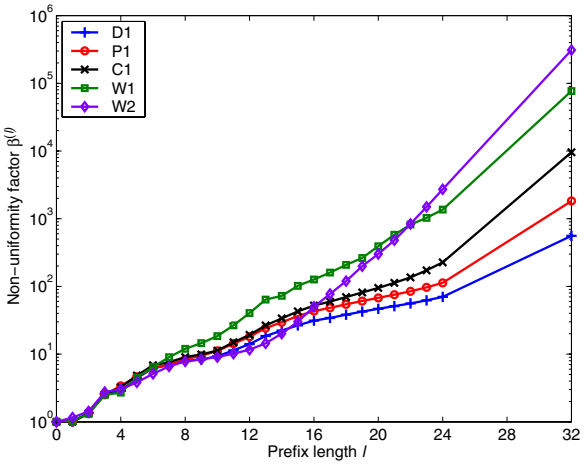
It is well-known that the Shannon entropy can be used to measure the non-uniformity of a distribution [8]. Why do we choose the non-uniformity factor instead?

Consider a general entropy, called the *Renyi entropy* [17], which is defined as

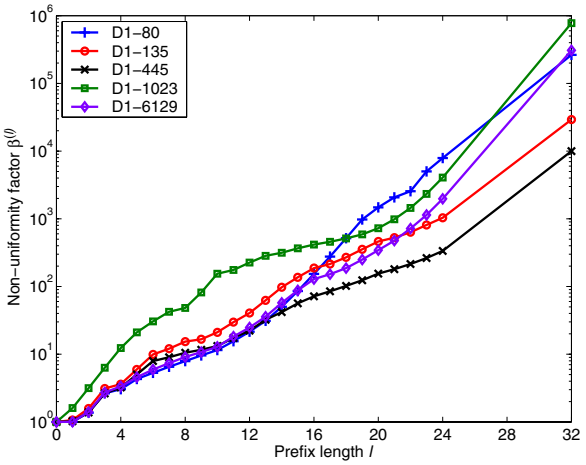
$$H_q\left(P^{(l)}\right) = \frac{1}{1-q} \log_2 \sum_{i=1}^{2^l} \left(p_g^{(l)}(i) \right)^q, \text{ for } q \neq 1, \quad (7)$$

where $P^{(l)} = \{p_g^{(l)}(1), p_g^{(l)}(2), \cdots, p_g^{(l)}(2^l)\}$. The non-uniformity factor can relate to the Renyi entropy of order two in the following equation:

$$\beta^{(l)} = 2^{l - H_2\left(P^{(l)}\right)}. \quad (8)$$



(a) Five data sets.



(b) D1 with different ports.

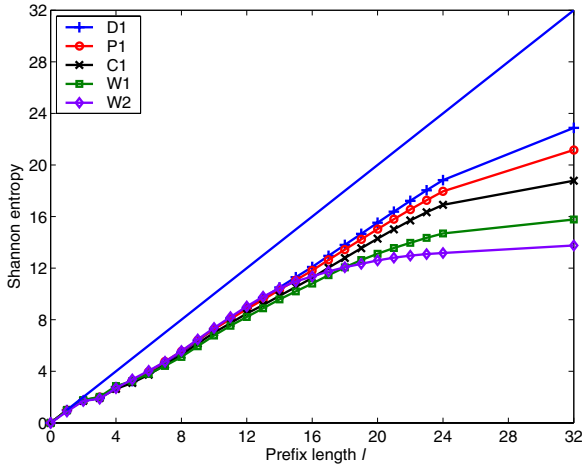
Fig. 2. Non-uniformity factors of collected data sets. The y-axis uses a *log* scale.

Thus, the non-uniformity factor is essentially an entropy.

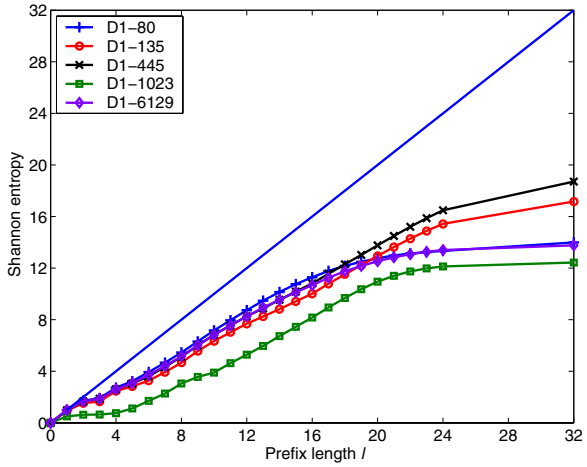
The Shannon entropy, $H\left(P^{(l)}\right) = -\sum_{i=1}^{2^l} p_g^{(l)}(i) \log_2 p_g^{(l)}(i)$, is a special case of the Renyi entropy [17], i.e.,

$$H\left(P^{(l)}\right) = \lim_{q \to 1} H_q\left(P^{(l)}\right). \quad (9)$$

Figure 3 shows Shannon entropies of our empirical distributions from the data sets. If a distribution is uniform, $H\left(P^{(l)}\right) = l$ as denoted by the diagonal line in the figure. On the other hand, if a distribution is extremely non-uniform, e.g., all vulnerable hosts concentrate on one subnet, $H\left(P^{(l)}\right) = 0$. Hence, the distance between $H\left(P^{(l)}\right)$ and 0 in Figure 3 reflects how uniform a distribution is. Similarly, the distance between $\beta^{(l)}$ and the horizontal access 1 in Figure 2 measures the degree of unevenness. A larger $H\left(P^{(l)}\right)$ corresponds to a more even distribution, whereas a larger $\beta^{(l)}$ corresponds to a more non-uniform distribution. Evidenced by Figure 2, the non-uniformity factor magnifies the unevenness of a distribution. In addition, if two distributions have different prefix lengths, we can directly apply the non-uniformity factor to compare the unevenness between them. Therefore, the non-



(a) Five data sets.



(b) D1 with different ports.

Fig. 3. Shannon entropies of collected data sets.

uniformity factor provides a better measure for describing the non-uniformity of a distribution.

More importantly, the non-uniformity factor can directly reflect how much faster a network-aware worm spreads than a random-scanning worm, which is shown in the next section.

From an information theoretical viewpoint, the entropy provides a quantitative measure of uncertainty. The uncertainty of a vulnerable-host probability distribution is important for an attacker to design an intelligent network-aware worm. If there is no uncertainty about the distribution of vulnerable hosts (e.g., either all vulnerable hosts are concentrated on a subnet or all information about vulnerable hosts is known), the entropy is minimum, and the worm that uses the information on the distribution can spread fastest by employing the optimal importance scanning [6]. On the other hand, if there is maximum uncertainty (e.g., vulnerable hosts are uniformly distributed), the entropy is maximum. But the worm cannot take advantage of the information of the distribution and can only use random scanning. Moreover, when an attacker obtains more information about the vulnerable-host distribution, in general, the resulting worm can spread faster.

V. Network-Aware Worm Spreading Ability

How to quantify the spreading speed of a network-aware worm with the information of a vulnerable-host distribution? We characterize the spread of a network-aware worm at an early stage by deriving the infection rate.

A. Infection Rate

The infection rate, denoted by α, is defined as the average number of vulnerable hosts that can be infected per unit time by one infected host during the early stage of worm propagation [26]. The infection rate is an important metric for studying network-aware worm spreading ability for two reasons. First, since the number of infected hosts increases exponentially with the rate $1+\alpha$ during the early stage, a worm with a higher infection rate can spread much faster at the beginning and thus infect a large number of hosts in a shorter time [6]. Second, while it is generally difficult to derive a close-form solution for dynamic worm propagation, we can obtain a close-form expression of the infection rate for different worm scanning methods.

Let R denote the (random) number of vulnerable hosts that can be infected per unit time by one infected host during the early stage of worm propagation. The infection rate is the expected value of R, i.e., $\alpha = E[R]$. Let s be the scanning rate or the number of scans sent by an infected host per unit time, N be the number of vulnerable hosts, and Ω be the scanning space (i.e., $\Omega = 2^{32}$).

For random scanning (RS) [26], [6], an infected host sends out s random scans per unit time, and the probability that one scan hits a vulnerable host is $\frac{N}{\Omega}$. Thus, R follows a Binomial distribution B(s, $\frac{N}{\Omega}$)[2], resulting in

$$\alpha_{RS} = E[R] = \frac{sN}{\Omega}. \tag{10}$$

B. Importance Scanning

We derive the infection rates of importance scanning (IS) [6], [5]. An infected host scans /l subnet i with the probability $q_g^{(l)}(i)$. $q_g^{(l)}(i)$ is called the group scanning distribution and is to be chosen with respect to the group distribution $p_g^{(l)}(i)$. If a worm scan hits /l subnet i, it would have a probability of $\frac{Np_g^{(l)}(i)}{2^{32-l}}$ to find a vulnerable host. Thus, a worm scan hits a vulnerable host with a likelihood of $\sum_{i=1}^{2^l}\left(q_g^{(l)}(i)\cdot\frac{Np_g^{(l)}(i)}{2^{32-l}}\right)$. Similar to random scanning, R of IS follows a Binomial distribution B(s, $\sum_{i=1}^{2^l}\frac{Np_g^{(l)}(i)q_g^{(l)}(i)}{2^{32-l}}$), which leads to

$$\alpha_{IS} = E[R] = sN\sum_{i=1}^{2^l}\frac{p_g^{(l)}(i)q_g^{(l)}(i)}{2^{32-l}}. \tag{11}$$

The same result was derived in [6] but by a different approach.

We now consider a special case of IS, where the group scanning distribution $q_g^{(l)}(i)$ is chosen to be proportional to the number of vulnerable hosts in group i, i.e., $q_g^{(l)}(i) = p_g^{(l)}(i)$.

[2]In our derivation, we ignore the dependency of the events that different scans hit the same target at the early stage of worm propagation.

This results in suboptimal IS [6], called /l IS. Thus, the infection rate is

$$\alpha_{IS}^{(l)} = \frac{sN}{2^{32-l}} \sum_{i=1}^{2^l} (p_g(i))^2 = \alpha_{RS} \cdot \beta^{(l)}. \quad (12)$$

Compared with RS, this /l IS can increase the infection rate by a factor of $\beta^{(l)}$. Such an infection rate can be considered as a benchmark for comparison with other network-aware worms.

C. Localized Scanning

Localized scanning (LS) has been used by such real worms as Code Red II and Nimda [15], [4]. We first consider a simplified version of LS, called /l LS, which scans the Internet as follows:

- p_a ($0 \le p_a \le 1$) of the time, an address with the same first l bits is chosen as the target,
- $1 - p_a$ of the time, a random address is chosen.

Assume that an initially infected host is randomly chosen from the vulnerable hosts. Let I_g denote the subnet where an initially infected host locates. Thus, $P(I_g = i) = p_g^{(l)}(i)$, where $i = 1, 2, \cdots, 2^l$. For an infected host located in /l subnet i, a scan from this host probes globally with the probability of $1 - p_a$ and hits /l subnet j ($j \neq i$) with the likelihood of $\frac{1-p_a}{2^l}$. Thus, the group scanning distribution for this host is

$$q_g^{(l)}(j) = \begin{cases} p_a + \frac{1-p_a}{2^l}, & \text{if } j = i; \\ \frac{1-p_a}{2^l}, & \text{otherwise,} \end{cases} \quad (13)$$

where $j = 1, 2, \cdots, 2^l$. Given the subnet location of an initially infected host, we can apply the results of IS. Specifically, putting Equation (13) into Equation (11), we have

$$E[R|I_g = i] = \frac{sN}{2^{32-l}} \left(p_a p_g^{(l)}(i) + \frac{1 - p_a}{2^l} \right). \quad (14)$$

Therefore, we can compute the infection rate of /l LS as

$$\alpha_{LS}^{(l)} = E[R] = E[E[R|I_g]] = \sum_{i=1}^{2^l} p_g^{(l)}(i) E[R|I_g = i], \quad (15)$$

resulting in

$$\alpha_{LS}^{(l)} = \alpha_{RS} \left(1 - p_a + p_a \beta^{(l)} \right). \quad (16)$$

Since $\beta^{(l)} > 1$ ($\beta^{(l)} = 1$ is for a uniform distribution and is excluded here), $\alpha_{LS}^{(l)}$ increases with respect to p_a. Specifically, when $p_a \to 1$, $\alpha_{LS}^{(l)} \to \alpha_{RS}\beta^{(l)} = \alpha_{IS}^{(l)}$. Thus, /$l$ LS has an infection rate comparable to that of /l IS. In reality, p_a cannot be 1. This is because an LS worm begins spreading from one infected host that is specifically in a subnet; and if $p_a = 1$, the worm can never spread out of this subnet. Therefore, we expect that the optimal value of p_a should be large but not 1.

Next, we further consider another LS, called two-level LS (2LLS), which has been used by the Code Red II and Nimda worms [28], [29]. 2LLS scans the Internet as follows:

- p_b ($0 \le p_b \le 1$) of the time, an address with the same first byte is chosen as the target,
- p_c ($0 \le p_c \le 1 - p_b$) of the time, an address with the same first two bytes is chosen as the target,
- $1 - p_b - p_c$ of the time, a random address is chosen.

For example, for the Code Red II worm, $p_b = 0.5$ and $p_c = 0.375$ [28]; for the Nimda worm, $p_b = 0.25$ and $p_c = 0.5$ [29]. Using the similar analysis for /l LS, we can derive the infection rate of 2LLS:

$$\alpha_{2LLS} = \alpha_{RS} \left(1 - p_b - p_c + p_b \beta^{(8)} + p_c \beta^{(16)} \right). \quad (17)$$

Since $\beta^{(16)} \ge \beta^{(8)} \ge 1$ from Theorem 1, α_{2LLS} holds or increases when both p_b and p_c increase. Specially, when $p_c \to 1$, $\alpha_{2LLS} \to \alpha_{RS}\beta^{(16)} = \alpha_{IS}^{(16)}$. Thus, 2LLS has an infection rate comparable to that of /16 IS. Moreover, $\beta^{(16)}$ is much larger than $\beta^{(8)}$ as shown in Table II for the collected distributions. Hence, p_c is more significant than p_b for 2LLS.

D. Modified Sequential Scanning

The Blaster worm is a real worm that exploits sequential scanning in combination with localized scanning. A *sequential-scanning* worm studied in [27], [10] begins to scan addresses sequentially from a randomly chosen starting IP address and has a similar propagation speed as a random-scanning worm. The Blaster worm selects its starting point locally as the first address of its Class C subnet with probability 0.4 [31], [27]. To analyze the effect of sequential scanning, we do not incorporate localized scanning. Specifically, we consider our /l modified sequential-scanning (MSS) worm, which scans the Internet as follows:

- Newly infected host A begins with random scanning until finding a vulnerable host with address B.
- After infecting the target B, host A continues to sequentially scan IP addresses $B + 1$, $B + 2$, $\cdots$ (or $B - 1$, $B - 2$, $\cdots$) in the /l subnet where B locates.

Such a sequential worm-scanning strategy is in a similar spirit to the *nearest neighbor rule*, which is widely used in pattern classification [7]. The basic idea is that if the vulnerable hosts are clustered, the neighbor of a vulnerable host is likely to be vulnerable also.

Such a /l MSS worm has two stages. In the first stage (called MSS_1), the worm uses random scanning and has an infection rate of α_{RS}, i.e., $\alpha_{MSS\_1} = \alpha_{RS}$. In the second stage (called MSS_2), the worm scans sequentially in a /l subnet. The fist l bits of a target address are fixed, whereas the last $32-l$ bits of the address are generated additively or subtractively and are modulated by 2^{32-l}. Let I_g denote the sunbet where B locates. Thus, $P(I_g = i) = p_g^{(l)}(i)$, where $i = 1, 2, \cdots, 2^l$. Since a sequential worm scan in subnet i has a probability of $\frac{N_i^{(l)}}{2^{32-l}}$ to hit a vulnerable host, $E[R|I_g = i] = \frac{N_i^{(l)}}{2^{32-l}} s = \alpha_{RS} \cdot 2^l p_g^{(l)}(i)$, which leads to

$$\alpha_{MSS\_2} = E[R] = E[E[R|I_g]] = \alpha_{RS} \cdot \beta^{(l)}. \quad (18)$$

Therefore, the infection rate of /l MSS is between α_{RS} and $\alpha_{RS}\beta^{(l)}$.

In Summary, infection rates of all three network-aware worms (IS, LS, and MSS) can be far larger than that of an RS worm, depending on the non-uniformity factors.

VI. Simulation and Validation

A. Infection Rate

We first focus on validating infection rates. We apply the discrete event simulation to our experiments [18]. Specifically, we simulate the searching process of a worm using different scanning methods at the early stage. We use the C1 data set for the vulnerable-host distribution. The worm spreads over the C1 distribution with $N = 448,894$ and has a scanning rate $s = 100$. Note that the C1 distribution has the non-uniformity factors $\beta^{(8)} = 9.0$ and $\beta^{(16)} = 52.2$. The simulation stops when the worm has sent out 10^3 scans for RS, /16 IS, /16 LS, and 2LLS, and 65,535 scans for /16 MSS_2. Then, we count the number of vulnerable hosts hit by the worm scans and compute the infection rate. The results are averaged over 10^4 runs. Table III compares the simulation results (i.e., sample mean) with the analytical results (i.e., Equations (10), (12), (16), (17), and (18)). Here, a /16 LS worm uses $p_a = 0.75$, whereas a 2LLS worm employs $p_b = 0.25$ and $p_c = 0.5$. We observe that the sample means and the analytical results are almost identical.

TABLE III
Infection rates of different scanning methods.

Scanning method	RS	/16 IS	/16 LS	2LLS	/16 MSS_2
Analytical result	0.0105	0.5456	0.4118	0.2989	0.5456
Sample mean	0.0103	0.5454	0.4023	0.2942	0.5489
Sample variance	0.0010	0.0543	0.2072	0.1053	0.3186

We observe that network-aware worms have much larger infection rates than random-scanning worms. LS indeed increases the infection rate with nearly a non-uniformity factor and approaches the capacity of suboptimal IS. This is significant as LS only depends on one or two parameters (i.e., p_a for /l LS and p_b, p_c for 2LLS), while IS requires the information of the vulnerable-host distribution. On the other hand, LS has a larger sample variance than IS as indicated by Table III. This implies that the infection speed of an LS worm depends on the location of initially infected hosts. If the LS worm begins spreading from a subnet containing densely populated vulnerable hosts, the worm would spread rapidly. Furthermore, we notice that the MSS worm also has a large infection rate at the second stage, indicating that MSS can indeed exploit the clustering pattern of the distribution. Meanwhile, the large sample variance of the infection rate of MSS_2 reflects that an MSS worm strongly depends on the initially infected hosts. We further compute the infection rate of a /16 MSS worm that includes both random-scanning and sequential-scanning stages. Simulation results are averaged over 10^6 runs and are summarized in Table IV. These results strongly depend on the total number of worm scans. When the number of worm scans is small, an MSS worm behaves similar to a random-scanning worm. When the number of worm scans increases, the MSS worm spends more scans on the second stage and thus has a larger infection rate.

B. Dynamic Worm Propagation

An infection rate only characterizes the early stage of worm propagation. We now employ the analytical active worm propagation (AAWP) model and its extensions to characterize the entire spreading process of worms [4]. Specifically, the spread of RS and IS worms is implemented as described in [6], whereas the propagation of LS worms is modeled according to [15]. The parameters that we use to simulate a worm are comparable to those of the Code Red v2 worm. Code Red v2 has a vulnerable population $N = 360,000$ and a scanning rate $s = 358$ per minute [25]. We assume that the worm begins spreading from an initially infected host that is located in the subnet containing the largest number of vulnerable hosts.

TABLE IV
Infection rates of a /16 MSS worm.

# of worm scans	10	100	1000	10000	50000
Sample mean	0.0108	0.0190	0.0728	0.2866	0.4298
Sample variance	0.1246	0.1346	0.1659	0.2498	0.2311

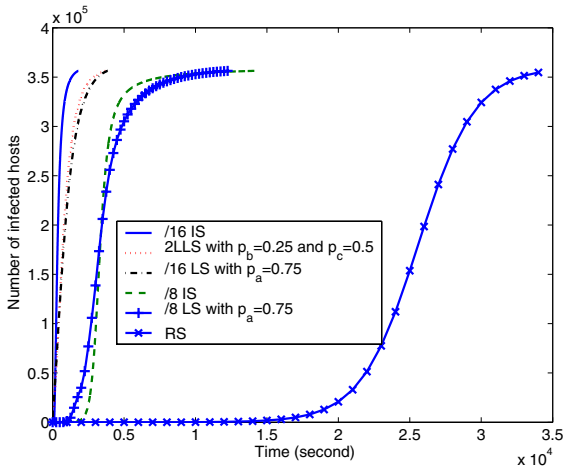


Fig. 4. A network-aware worm spreads over the D1-80 distribution.

We first show the propagation speeds of network-aware worms for the same vulnerable-host distribution from data set D1-80. From Section V, we expect that a network-aware worm can spread much faster than an RS worm. Figure 4 demonstrates such an example on a worm that uses different scanning methods. It takes an RS worm 10 hours to infect 99% of vulnerable hosts, whereas a /8 LS worm with $p_a = 0.75$ or a /8 IS worm takes only about 3.5 hours. A /16 LS worm with $p_a = 0.75$ or a 2LLS worm with $p_b = 0.25$ and $p_c = 0.5$ can further reduce the time to 1 hour. A /16 IS worm spreads fastest and takes only 0.5 hour.

We also study the effect of vulnerable-host distributions on the propagation of network-aware worms. From Table II, we observe that $\beta^{(16)}_{D1-1023} > \beta^{(16)}_{W1} > \beta^{(16)}_{C1} > \beta^{(16)}_{D1}$. Thus, we expect that a network-aware worm using the /16 D1-1023 distribution would spread faster than using other three distributions. Figure 5 verifies this through the simulations of the spread of a 2LLS worm that uses different vulnerable-host distributions (i.e., D1-1023, W1, C1, and D1). Here, the 2LLS worm employs the same parameters as the Nimda worm, i.e., $p_b = 0.25$ and $p_c = 0.5$. As expected, the worm using the D1-1023 distribution spreads fastest, especially at the early stage of worm propagation.

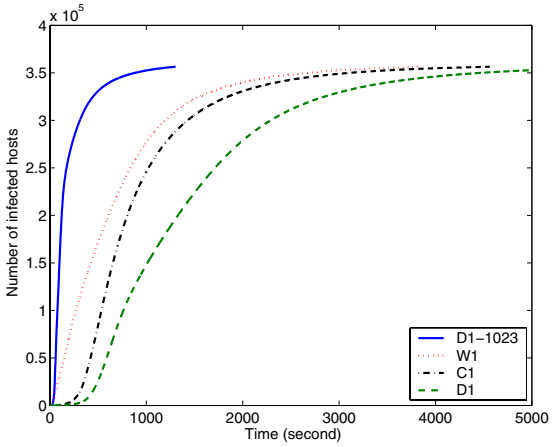


Fig. 5. A 2LLS worm spreads over different distributions.

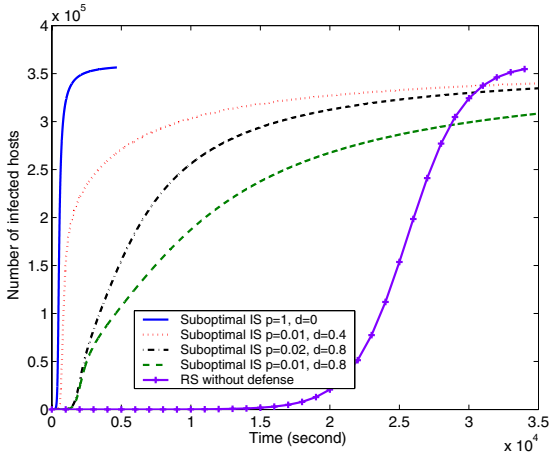


Fig. 6. A /16 IS worm spreads under the defense of PP.

VII. Effectiveness of Defense Strategies

A. Host-Based Defense

Host-based defense has been widely used for random-scanning worms. Proactive protection and virus throttling are examples of host-based defense strategies.

A *proactive protection* (PP) strategy proactively hardens a system, making it difficult for a worm to exploit vulnerabilities [3]. Techniques used by PP include address-space randomization, pointer encryption, instruction-set randomization, and password protection. Thus, a worm requires multiple trials to compromise a host that implements PP. Specifically, let p ($0 \leq p \leq 1$) denote the protection probability or the probability that a single worm attempt succeeds in infecting a vulnerable host that implements PP. On the average, a worm should make $\frac{1}{p}$ exploit attempts to compromise the target. We assume that hosts with PP are uniformly deployed in the Internet. Let d ($0 < d \leq 1$) denote the deployment ratio between the number of hosts with PP and the total number of hosts.

To show the effectiveness of the PP strategy, we consider the infection rate of a /l IS worm. Since now some of the vulnerable hosts implement PP, Equation (12) changes to

$$\begin{aligned} \alpha_{IS}^{(l)} &= \frac{sN}{2^{32-l}} \sum_{i=1}^{2^l} \left[dp \left(p_g^{(l)}(i) \right)^2 + (1-d) \left(p_g^{(l)}(i) \right)^2 \right] \\ &= \alpha_{RS} \beta^{(l)} (1 - d + dp). \end{aligned} \tag{19}$$

To slow down the spread of a suboptimal IS worm to that of a random-scanning worm, $\beta^{(l)}(1-d+dp) \leq 1$, resulting in

$$p \leq \frac{1 - (1-d)\beta^{(l)}}{d\beta^{(l)}}. \tag{20}$$

When PP is fully deployed, i.e., $d = 1$, p can be at most $\frac{1}{\beta^{(l)}}$. On the other hand, if PP provides perfect protection, i.e., $p = 0$, d should be at least $1 - \frac{1}{\beta^{(l)}}$. Therefore, when $\beta^{(l)}$ is large, Inequality (20) presents high requirements for the PP strategy. For example, if $\beta^{(16)} = 50$ (most of $\beta^{(16)}$'s in Table II are larger than this value), $p \leq 0.02$ and $d \geq 0.98$. That is, a PP strategy should be almost fully deployed and provide a nearly perfect protection for a vulnerable host.

We extend the model described in [6] to characterize the spread of suboptimal IS worms under the defense of the PP strategy and show the results in Figure 6. Here, Code-Red-v2-like worms spread over the C1 distribution with $\beta^{(16)} = 52.2$. It is observed that even when the protection probability is small (e.g., $p = 0.01$) and the deployment ratio is high (e.g., $d = 0.8$), a /16 IS worm is slowed down a little at the early stage, compared with a /16 IS worm without the PP defense (i.e., $p = 1$ and $d = 0$). Moreover, when p is small (e.g., $p \leq 0.02$), d is a more sensitive parameter than p.

We next consider the *virus throttling* (VT) strategy that constrains the number of outgoing connections of a host [22]. Thus, VT can reduce the scanning rate of an infected host. We find that Equation (19) also holds for this strategy, except that p is the ratio between the scanning rate of infected hosts with VT and that of infected hosts without VT. Therefore, VT also requires to be almost fully deployed for fighting network-aware worms effectively.

From these two strategies, we have learned that an effective strategy should reduce either α_{RS} or $\beta^{(l)}$. Host-based defense, however, is limited in such capabilities shown in this section.

B. IPv6

IPv6 can decrease α_{RS} significantly [26] by increasing the scanning space. But the non-uniformity factor would increase the infection rate if the vulnerable-host distribution is still non-uniform. Hence, an important question is whether IPv6 can counteract network-aware worms when both α_{RS} and $\beta^{(l)}$ are taken into consideration.

We study this issue by computing the infection rate of a network-aware worm in the IPv6 Internet. As pointed out by [2], a smart worm can first detect some vulnerable hosts in /64 subnets containing many vulnerable hosts, then release to the hosts on the hitlist, and finally spread inside these subnets. Such a worm only scans the local /64 subnet. Thus, we focus on the spreading speed of a network-aware worm in a /64 subnet. From Figure 2, we extrapolate that $\beta^{(32)}$ in the IPv6 Internet can be in the order of 10^5 if hosts are still distributed in a clustered fashion. Using the parameters $N = 10^8$ proposed by [9] and $s = 4,000$ used by the

Slammer worm [13], we derive the infection rate of a /32 IS worm in a /64 subnet of the IPv6 Internet: $\alpha_{IS}^{IPv6} = \frac{sN}{2^{64}} \cdot \beta^{(32)} = 2.2 \times 10^{-3}$. α_{IS}^{IPv6} is larger than the infection rate of the Code Red v2 worm in the IPv4 Internet, where $\alpha_{RS}^{CR} = \frac{360{,}000 \times 358/60}{2^{32}} = 5 \times 10^{-4}$.

Therefore, IPv6 can only slow down the spread of a network-aware worm to that of a random-scanning worm in IPv4. To defend against the worm effectively, we should further consider how to slow down the increase rate of $\beta^{(l)}$ as l increases when IPv4 is updated to IPv6.

VIII. Conclusions

In this paper, we have observed and characterized non-uniform vulnerable-host distributions across five measurement sets from different sources. We have derived a simple metric, known as the non-uniformity factor, to quantify an uneven distribution of vulnerable hosts. The non-uniformity factors have been obtained using our collected data, and all of which demonstrate large values. This implies that the non-uniformity of the vulnerable-host distribution is significant and seems to be consistent across networks and applications. Moreover, the non-uniformity factor, shown as a function of the Renyi entropy of order two, better characterizes the uneven feature of a distribution than the Shannon entropy.

The importance of a non-uniformity factor is that it quantifies the spreading ability of network-aware worms. We have derived analytical expressions relating the non-uniformity factors with the infection rates of network-aware worms. We have empirically verified that localized scanning and modified sequential scanning can increase the infection rate by nearly a non-uniformity factor when compared to random scanning and thus approach the capacity of suboptimal importance scanning.

Furthermore, we have evaluated the effectiveness of several commonly used defense strategies on network-aware worms. The host-based defense, such as proactive protection or virus throttle, requires to be almost fully deployed to slow down worm spreading at the early stage. This implies that host-based defense would be weakened significantly by network-aware scanning. More surprisingly, different from previous findings, we have shown that network-aware worms can be zero-day worms in the IPv6 Internet if vulnerable hosts are still clustered. These findings present a significant challenge to worm defense: Entirely different strategies may be needed for fighting against network-aware worms.

As part of our ongoing work, we plan to develop effective detection and defense systems against network-aware worms, taking the vulnerable-host distribution into consideration.

References

[1] P. Barford, R. Nowak, R. Willett, and V. Yegneswaran, "Toward a model for sources of Internet background radiation," in *Proc. of the Passive and Active Measurement Conference (PAM'06)*, Mar. 2006.

[2] S. M. Bellovin, B. Cheswick, and A. Keromytis, "Worm propagation strategies in an IPv6 Internet," *;login:*, vol. 31, no. 1, Feb. 2006, pp. 70-76.

[3] D. Brumley, L. Liu, P. Poosankam, and D. Song, "Design space and analysis of worm defense strategies," in *ACM Symposium on InformAtion, Computer and Communications Security (ASIACCS)*, Mar. 2006.

[4] 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.

[5] 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.

[6] Z. Chen and C. Ji, "Optimal worm-scanning method using vulnerable-host distributions," *to appear in the International Journal of Security and Networks: Special Issue on Computer and Network Security*, 2007.

[7] T. M. Cover and P. E. Hart, "Nearest neighbor pattern classification," *IEEE Transactions on Information Theory*, vol. IT-13, no. 1, Jan. 1967.

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

[9] 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.

[10] G. Gu, M. Sharif, X. Qin, D. Dagon, W. Lee, and G. Riley, "Worm detection, early warning and response based on local victim information," in *Proc. 20th Ann. Computer Security Applications Conf. (ACSAC'04)*, Tucson, AZ, Dec. 2004.

[11] E. Kohler, J. Li, V. Paxson, and S. Shenker, "Observed structure of addresses in IP traffic," in *ACM SIGCOMM Internet Measurement Workshop*, Marseille, France, Nov. 2002.

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

[13] 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.

[14] Y. Pryadkin, R. Lindell, J. Bannister, and R. Govindan, "An empirical evaluation of IP address space occupancy," *Technical Report ISI-TR-2004-598*, USC/Information Sciences Institute, Nov. 2004.

[15] 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.

[16] M. A. Rajab, F. Monrose, and A. Terzis, "Fast and evasive attacks: highlighting the challenges ahead," in *Proc. of the 9th International Symposium on Recent Advances in Intrusion Detection (RAID'06)*, Hamburg, Germany, Sept. 2006.

[17] A. Renyi, *Probability Theory*. North-Holland, Amsterdam, 1970.

[18] S. M. Ross, *Simulation*, 3rd Edition. Academic Press, 2002.

[19] 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.

[20] 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.

[21] 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.

[22] J. Twycross and M. M. Williamson, "Implementing and testing a virus throttle," in *Proc. of the 12th USENIX Security Symposium (Security'03)*, Washington, DC, Aug. 2003, pp. 285-294.

[23] 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.

[24] V. Yegneswaran, P. Barford, and D. Plonka, "On the design and utility of Internet sinks for network abuse monitoring," in *Proc. of Symposium on Recent Advances in Intrusion Detection (RAID'04)*, 2004.

[25] 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.

[26] 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.

[27] 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.

[28] CERT Coordination Center, "'Code Red II:' another worm exploiting buffer overflow in IIS indexing service DLL," CERT Incident Note IN-2001-09, http://www.cert.org/incident_notes/IN-2001-09.html.

[29] CERT Coordination Center, CERT Advisory CA-2001-26 Nimda Worm, http://www.cert.org/advisories/CA-2001-26.html.

[30] Distributed Intrusion Detection System (DShield), http://www.dshield.org/.

[31] eEye Digital Security, "ANALYSIS: Blaster worm," http://www.eeye.com/html/Research/Advisories/AL20030811.html.

[32] Internet Protocol V4 Address Space, http://www.iana.org/assignments/ipv4-address-space.