ROMAI J., **1**, 2(2005),
133–142

# INTERNET WORMS: PROPAGATION MODELING AND ANALYSIS

Victor-Valeriu Patriciu, Iustin Priescu, Sebastian Nicolaescu
*Department of Computer Engineering, Military Technical Academy,*
*Department of Computer Engineering, Military Technical Academy,*
*MCI, USA*
vip@mta.ro, iustin@mta.ro, sebastian.nicolaescu@mci.com

**Abstract** Several Internet-scale incidents from the recent years have demonstrated the ability of self-propagating code, also known as "network worms", to infect large numbers of hosts, exploiting vulnerabilities in the largely deployed operating systems and applications. Capable of infecting a substantial portion of the hosts within several minutes, and impacting the world-wide network operations by generating a distributed denial of service (DDoS) attack on the whole Internet, network worms are considered a major security threat. So, a better understanding of the worms' propagation means will help to implement a more efficient detection and defense.

## 1. INTRODUCTION

Recent worm incidents have indicated a lot of interest in implementing a variety of scanning strategies to increase the worms' spreading speed and defeat security defense measures. In this paper, we present some mathematical models to analyze various scanning strategies that attackers have already used or might use in the future. Mathematical analysis provides a better understanding of how multiple factors affect a worm's propagation, and it can help us to build a better defense against future worms. The scanning strategies presented in the paper include idealized scan, uniform scan, divide-et-impera scan, local preference scan, sequential scan.

## 2. EPIDEMIC MODELS

Computer worms look similar to biological viruses in their propagation behaviors and self-replication. Thus, the mathematical models developed for the study of biological infectious diseases can be adapted to the study of computer worm propagation. We briefly introduce two classical deterministic epidemic models: simple epidemic model in homogeneous system and in interacting groups, respectively. The models and analyses presented in this paper are based primarily on these two models and their underlying principles.

A. *Simple epidemic model for homogeneous systems*

The simple epidemic model is characterized by the following assumptions:

- each and every host occupies one of two states: susceptible or infectious (this model is also named in the literature as „SI model");
- once a host is infected, it stays permanently in the infectious state;
- each host is assumed to have equal probability to contact any other host.

The equation of the model is:

$$\frac{dI(t)}{dt} = \beta I(t)[N - I(t)], \tag{1}$$

where $N$ is the total number of hosts under consideration, $I(t)$ is the total number of infectious hosts at time $t$, $\eta$ is the average worm scan rate, $\Omega$ is the size of the worm's scanning space, $\beta$ is the pairwise rate of infection in worm propagation model, $\beta = \eta/\Omega$.

At $t = 0$, there are $I(0)$ infected hosts, and $[N - I(0)]$ susceptible. The epidemic model SI has the following analytical solution

$$I(t) = \frac{I(0)N}{I(0) + [N - I(0)]e^{-\beta N t}}, \tag{2}$$

The spreading of an Internet worm or an epidemic disease is, in fact, a stochastic process, but when considering a large-scale system consisting of a large population $N$, which is the case of an Internet worm, it can be used a mean value analysis based on the law of large numbers. Also, in the Internet context, an infected host has equal probability of contacting any other host when the worm uniformly scans the Internet. Thus, a uniform scan worm can be modeled the same way as an epidemic disease in a homogeneous system.

B. *Simple epidemic model for interacting groups*

It is an extension of SI model, defined by equation (1), for non-homogeneous systems. In this model, the system consists of $K$ groups; each group has population $N_1$, $N_2$, ..., $N_k$. Interactions between groups are different from interactions within a group. The equation of the model becomes

$$\frac{dI_k(t)}{dt} = [\beta_{kk} I_k(t) + \Sigma_{j \neq k} \beta_{jk} I_j(t)][N_k - I_k(t)], \tag{3}$$

where $k \in \{1, ..., K\}$, $\beta_{jk}$ is the infection rate, per interacting pair, of the susceptible hosts in the $k$-th group by the infectious hosts in the j-th group.

# 3. THE MODELS AND THE ANALYSIS OF WORM SCANNING STRATEGIES

3.1. *Idealized worm*

This kind of worms would posses the IP addresses of all vulnerable hosts in the Internet. Even if it is virtually impossible to implement them on the global scale of the Internet, this type of scanning strategy is valuable as a research model.

3.1.1. *Perfect worm*

Considered to be the fastest propagation worm, it knows the addresses of all vulnerable hosts in the Internet, and all infected hosts fully cooperate with each other such that they will not try to scan and infect an already infected host. For this type of worm, any scan is successful in infecting another vulnerable host.

The propagation model for the perfect worm is

$$\frac{dI(t)}{dt} = \begin{cases} \eta I(t), \ I(t) < N, \\ 0, \ I(t) = N. \end{cases} \tag{4}$$

Suppose the perfect worm starts with $I(0)$ infected hosts, the solution for (4) is

$$I(t) = min[I(0)e^{\eta t}, \ N]. \tag{5}$$

To illustrate the propagation speed of the perfect worm, we assume it has some of the parameters identified for Code Red ($\eta = 358$ hosts/minute, $N = 360000$ total vulnerable hosts, and $I(0) = 10$).

From (5) it follows that all vulnerable population will be infected in $T = [lnN - lnI(0)]/\eta = 1.76sec$. However, the above scenario did not consider various time delays in the worm's propagation like: the time to transfer the worm code to the vulnerable host, the time to execute the worm infectious code. If we consider the delay (e) from the moment when a worm scan is sent out, to the moment when the vulnerable host just infected begins to infect others, then the propagation model for the perfect worm becomes

$$\frac{dI(t)}{dt} = \begin{cases} \eta I(t-\varepsilon), \ I(t) < N, \\ 0, \ I(t) = N, \end{cases} \tag{6}$$

where $I(t-\varepsilon) = 0, \ \forall t < \varepsilon$.

3.1.2. *Flash worm*

Staniford et al. defined the „flash worm", as one which knows the IP addresses of all vulnerable hosts in the Internet ($N = \Omega$) and uniformly scans the vulnerable population. The propagation of a flash worm satisfies the epidemic

spreading assumptions in a homogeneous system and can be modeled by the simple epidemic model (1). The equation for the flash worm becomes

$$\frac{dI(t)}{dt} = \eta I(t)[N - I(t)]/N. \tag{7}$$

Assuming some of the parameters identified for Code Red ($\eta = 358$ hosts/minute, $N = 360000$ vulnerable hosts, and $I(0) = 10$), it can be determined that a flash worm can infect almost all the vulnerable host in $T = 2.5$ seconds. Based on these results, the delays due to code propagation are significant and should be factored in the worm's propagation model. Assuming a propagation delay $\varepsilon$, (7) becomes

$$\frac{dI(t)}{dt} = \eta I(t-\varepsilon)[N - I(t)]/N,\ textwhere\ I(t-\varepsilon) = 0,\ \forall t < \varepsilon. \tag{8}$$

3.2. *Uniform scan worms*

When a worm does not have the IP addresses of the vulnerable hosts, the simplest solution is to scan randomly the entire IPv4 space ($\Omega = 232$) in order to identify potential victims. This scanning strategy was used by Code Red (07/2001) and Slammer (01/2003). Based on (1) the propagation model for Code Red type worms becomes

$$\frac{dI(t)}{dt} = 2^{-32}\eta I(t)[N - I(t)]/N. \tag{9}$$

An analysis between (7) and (9) shows that the scanning space of a flash worm is much smaller than IPv4 (used by a Code Red type worm). Also, because the propagation delays (e) are negligible in comparison with worm's spreading speed, we can ignore them.

3.2.1. *"Hit List" worm*

This type of worm was defined by Staniford, as a way to improve the spreading speed of a uniform scan worm. It has a list with IP addresses of some vulnerable host in Internet. First, the worm behaves like a „flash worm" by scanning and infecting the hosts from hit-list, and then it scans randomly the entire IPv4 space to infect other vulnerable hosts.

3.2.2. *"Routing Worm"*

Based on the BGP routing table, it has been established that only 28.6% of the IPv4 space is routable. With no change in the scanning strategy, any worm can improve its performances by scanning a smaller IP address space. Many of the recent worms have already taken of this information. For example, the scanning space for Win32.Doomjuice.a (2004) consisted of 160 class A networks ($\Omega = 160 * 224$)

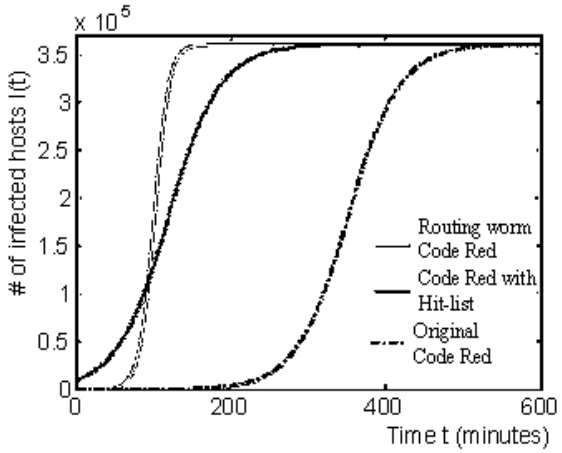


Fig 1. The Propagation Performances for possible Code Red versions taking advantage of "Hit List" and "Routing Worm" improvements.

It can be noticed that „Hit List" worm with $I(0) = 10.000$ can infect more hosts in a short time due to its hit-list, but it has a slower speed than a „Routing worm" with $\Omega = 0.286 * 2^{32}$.

3.2.3. *"Divide-et-impera" scan worm*

Another possibility to enhance the performances of a uniform scan worm would be to use a „divide-et-impera" approach to allow different infected hosts to scan and infect vulnerable hosts on distinct IP spaces. In the propagation of this kind of worm, there will be no case when two infected hosts would try to probe the same target.

The model assumes there is only one infected host initially in the system, and that vulnerable hosts are uniformly distributed in the entire scanning space O. Each infected host uniformly scans IP addresses in its scanning space. Once a target is infected, half of the scanning space allocated to the host that infected the target is transferred to the target (the space passed to the target includes the target host), while the infecting host remains with half of its original scanning space.

Based on these assumptions, none of the infected host will be subject to be probed, and the scanning space will be $\Omega' = \Omega - I(t)$. So, the equation for a uniform scan worm that uses a „Divide-et-impera" strategy is:

$$\frac{dI(t)}{dt} = \eta I(t)[N - I(t)]/(\Omega - I(t)), \tag{10}$$

For Internet worms, the number of vulnerable hosts $N$ is much smaller than $\Omega$ ($I(t) < N << \Omega$). Therefore, $\Omega - I(t) \approx \Omega$ which means that when vulnerable hosts are uniformly distributed, a „divide-et-impera" scan worm

propagates in the same way as a uniform scan worm, and can be modeled by the simple epidemic model (1).

3.3. *Subnet scan worm*

The uniform scan is the simplest scanning strategy that a worm may use. However, it is not the optimal one because the vulnerable hosts are not uniformly distributed in Internet. A worm could increase its spreading speed when it scans with a higher probability in the IP spaces that have a higher density of vulnerable hosts.

In the following, we model and analyze a subnet scanning worm that has probability p to uniformly scan IP addresses in its own "/n" prefix subnetwork and probability (1-p) to uniformly scan other IP addresses. A „/n" prefix subnetwork represents the IP space where the addresses have the same first n bits. Thus, in the current IPv4 Internet, a „/n" prefix subnetwork contains $2^{32-n}$ IP addresses.

This class of worms can be modeled based on the simple epidemic model for interacting groups (3), where the worm scanning space $\Omega$ consists of K „/n" prefix subnetworks ($\Omega = K * 2^{32-n}$), $\beta' = \beta_{kk}$ - represents the pairwise rate of infection in local scan and $\beta'' = \beta_{jk} = \beta_{kj}$ is the pairwise rate of infection in remote scan. The model's equation becomes

$$\frac{dI(t)}{dt} = [\beta' I_k(t) + \Sigma_{j \neq k} \beta'' I_j(t)][N_k - I_k(t)], \tag{11}$$

where $k \in \{1, \ ..., \ K\}$ and

$$\beta'' = \frac{p\eta}{2^{32-n}}, \ \beta'' = \frac{1-p)\eta}{(K-1)2^{32-n}}. \tag{12}$$

The type of analysis can be easily extended to other kinds of local preference scan strategies, such as local preference scanning with several levels of locality. For example, Code Red II (08/2001) had two-level locality in its local preference scan - the worm scanned the local class A network with a probability $p_A = 0.5$ and the local Class B network with a probability $p_B = 0.375$. Another example is Win32/Sasser.worm (05/2004) that probed the local class A network with a probability $p_A = 0.25$ and the local Class B network with a probability $p_B = 0.25$.

When vulnerable hosts are uniformly distributed in a worm's scanning space, and so the worm propagation in each subnetwork is identical, subnet scanning does not help a worm in its propagation speed.

Assuming the vulnerable hosts are uniformly distributed in $m$ out of the $K$ subnetworks ($N_1 = ... = N_m = N/m$) and all other $(K - m)$ subnetworks are not allocated ($N_{m+1} = ... = N_K = 0$), the equation for the each subnetwork

becomes

$$\frac{dI(t)}{dt} = [\beta' + (m-1)\beta'']I_k(t)][N_k - I_k(t)], \text{ where } k = 1, ..., m. \tag{13}$$

The equation for the propagation in Internet (for all subnetworks) is

$$\frac{dI(t)}{dt} = m\frac{dI_1(t)}{dt} = [\beta' + (m-1)\beta'']I(t)][N - I(t)]/m. \tag{14}$$

If a subnet scan worm wants to propagate as fast as possible, the worm should select a preference probability p that maximize the pairwise rate of infection $[\beta' + (m-1)\beta"]/m$ in (14). Thus, the optimal preference probability should be $p = 1$. Such a conclusion may seem unexpected, but it is reasonable based on the assumption used that all those $m$ „/n" subnetworks are considered identical. If $p = 1$, which means a worm only scans its own subnetwork, then no worm scans will be wasted in those $(K - m)$ empty subnetworks. In this way, the worm achieves its fastest spreading speed. In reality, no subnetwork is exactly the same as the others, and the worm has to scan remote networks in order to propagate to every part of the entire Internet.

The next chart shows the comparative simulation results between a Class A routing worm, subnet scan worm ($K = 256,\ m = 116$) for various local preference probabilities, and the original Code Red worm.

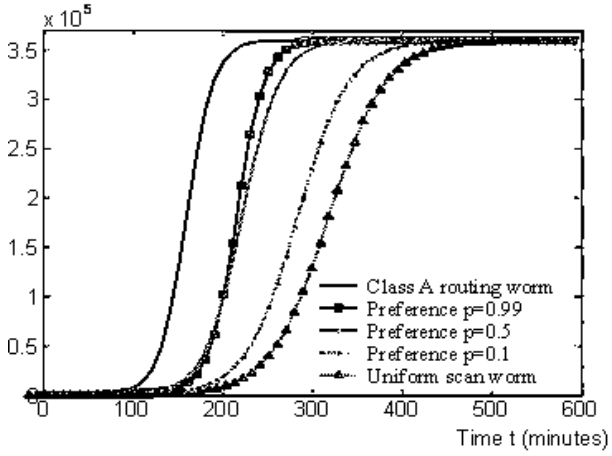


Fig. 2. Comparison of a Class A routing worm, subnet scan worm ($K = 256,\ m = 116$) and the original Code Red worm.

When vulnerable hosts are not uniformly distributed in a worm's scanning space, subnet scan increases the worm's propagation speed comparing with uniform scan. The optimal local preference scan probability $p$ increases when the local scan is on larger subnetworks.

3.4. *Sequential scan worm*

This scanning strategy assumes that once a vulnerable host is infected, it selects a starting address (x) from where it scans IP addresses sequentially (x+1, x+2,...).

A worm can select the start IP address (x) randomly, or close to its own address with higher probability. For example, for its starting point, the Blaster (08/2003) used the first address of the host's Class C subnetwork with probability $p = 0.4$, and chose a random IP address with a probability 0.6.

If vulnerable hosts are uniformly distributed in a worm's scanning space, a random sequential scan worm has the same propagation speed as a uniform scan worm and can be modeled by the SI epidemic model (1).

If a worm uses a local preference in selecting the start address, the worm would propagate slower as the child worm copies are more likely to be wasted on repeating their parents' scanning trails.

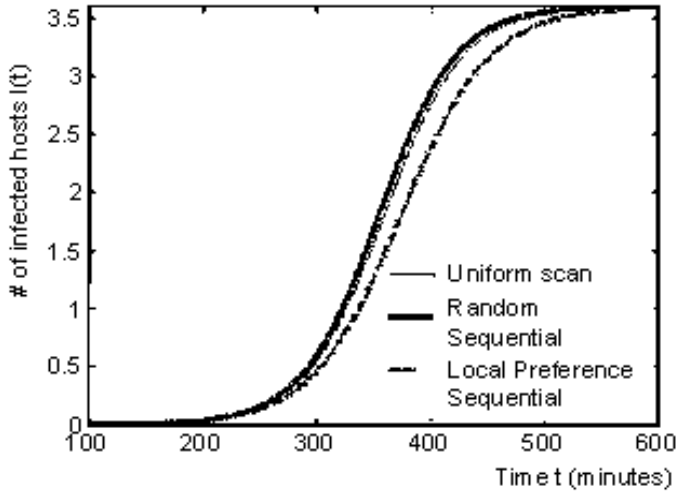


Figure 3. Comparison of a random sequential scan worm, a sequential scan worm with 0.4 local preference, and a uniform scan worm (100 simulation runs; vulnerable hosts uniformly distributed in entire IPv4 space).

A recent example of a sequential scan worm is Win32.Doomjuice.a (01/2004). It used a random sequential scan targeted to 160 Class A networks ($\Omega = \Omega 160 * 2^{24}$) [4].

## 4. CONCLUSIONS

In terms of the scanning strategies, the summary results of the analysis presented in this paper are:

- a subnet scan increases a worm's propagation speed when vulnerable hosts are not uniformly distributed. The optimal local preference probability increases when the subnet scan is on larger subnetworks;
- when vulnerable hosts are uniformly distributed, the divide-et-impera scan, the sequential scan, and the uniform scan are equivalent in terms of the total number of infected hosts at any time;
- for a sequential scan worm, using local preference in selecting the starting point slows down the worm's propagation speed.

Also, there are some results that should be considered in designing of a worm defense system:

- it is crucial to prevent attackers from identifying the IP addresses of a large number of vulnerable hosts ($I(0)$), or obtaining the address information that helps them to reduce significantly the worm's scanning space ($\Omega$);
- it is necessary to detect the worm as early as possible to be able to avoid a breakout and contain the propagation;
- a worm monitoring system should cover many well distributed IP blocks in order to accurately monitor the propagation of a non-uniform scan worm, especially a sequential scan worm such as Blaster.

## References

[1] Y. Moreno, R. Pastor-Satorras, A. Vespignani, *Epidemic outbreaks in complex heterogeneous networks*, The European Physical Journal B, **26**2002, 521-529.

[2] J. D. Murray, *Mathematical biology*, 2nd. corr. ed., Springer, New York, 1993.

[3] Z. Chen, L. Gao, K. Kwiat, *Modeling the spread of active worms*, IEEE INFOCOM, 2003.

[4] J. O. Kephart, D. M. Chess, S. R. White, *Computers and epidemiology*, IEEE Spectrum, 1993.

[5] J. O. Kephart, S. R. White, *Measuring and modeling computer virus prevalence*, in IEEE Symposimum on Security and Privacy, 1993.

[6] D. Moore, C. Shannon, J. Brown, *Code-Red: a case study on the spread and victims of an Internet worm*, in Proc. ACM/USENIX Internet Measurement Workshop, France, November, 2002.

[7] D. Moore, V. Paxson, S. Savage, C. Shannon, S. Staniford, N. Weaver, *Inside the Slammer worm*, IEEE Magazine on Security and Privacy, **1**4 (2003) 33-39.

[8] S. Staniford, V. Paxson, N. Weaver, *How to Own the Internet in your spare time*, in 11th Usenix Security Symposium, San Francisco, August, 2002.

[9] N. Weaver, V. Paxson, S. Staniford, and R. Cunningham, *A Taxonomy of Computer Worms*, ACM Workshop on Rapid Malcode, Washington, DC, Oct. 27, 2003.

[10] C.C. Zou, W. Gong, D. Towsley, *Code Red worm propagation modeling and analysis*, in 9th ACM Symposium on Computer and Communication Security, Washington DC, 2002.

[11] C.C. Zou, D. Towsley, W. Gong, S. Cai, *Routing Worm: a fast, selective attack worm based on IP Address Information*, Univ. Massachusetts Technical Report TRCSE- 03-06, November, 2003.