# Measurement and Analysis of Worm Propagation on Internet Network Topology

Jonghyun Kim, Sridhar Radhakrishnan, Sudarshan K. Dhall School of Computer Science, University of Oklahoma, Norman, Oklahoma, USA {jhkim, sridhar, <u>sdhall}@ou.edu</u>

## Abstract

There has been a constant barrage of worms over the internet during the recent past. Besides threatening network security, these worms cause an enormous economic burden in terms of loss of productivity at the victim hosts. In addition, these worms create unnecessary network data traffic that causes network congestion, thereby hurting all users. To develop appropriate tools for thwarting quick spread of worms, researchers are trying to understand the behavior of the worm propagation with the aid of epidemiological models. In this study, we apply the classical SIS model and a modification of SIR model to simulate worm propagation in two different network topologies. Whereas in the SIR model once a node is cured after infection it becomes permanently immune, our modification allows this immunity to be temporary, since the cured nodes may again become infected, maybe with a different strain of the same worm. The simulation study also shows that time to infect a large portion of the network vary significantly depending on where the infection begins. This information could be usefully employed to choose hosts for quarantine to delay worm propagation to the rest of the network.

# 1. Introduction

Lately, computer worms have become a major problem for the security of computer networks, causing considerable amount of resources and time to be spent recovering from virulent attacks. In general, worms, defined as selfpropagating codes, have been developed since the Morris worm arose in 1988 [4]. The convenience of Internet makes it more vulnerable for malicious Internet exploits. In other words, the Internet has become a powerful means for propagating malicious programs like computer viruses and worms. In the area of virus and worm modeling, many studies have employed simple epidemiological models to understand general characteristics of worm's propagation. Epidemiologic propagation models have traditionally been used to understand and model the spread of biological infectious diseases [2, 5]. A constant infection rate is reasonable for modeling epidemics but may not be valid for real Internet viruses and worms. The reason is that most classical epidemic models are homogeneous, in the sense that an infected host is equally likely to infect any of the susceptible hosts while Internet is non-homogeneous. In addition, current propagation studies *have not* considered the real Internet topology data and exploited characteristics of the network topology.

Previous works on worm modeling neglect the impacts of multiple worm outbreaks on our computer networks. Nowadays, new network worms will continue to be created while the strains of old worms will continue to circulate around the Internet. Recently, the Blaster worm, known as MSBlast or LoveSAN, has infected an estimated 188,000 systems running Microsoft operating systems that are unpatched for the so-called RPC vulnerability [13]. It is noted that 188,000 infected hosts is a substantial rate of infection, though the several hundred thousand hosts may be still infected by other old Internet worms including Slammer, Code Red and Nimda. Thus, any proposed defense mechanism must be evaluated in handling many active worms simultaneously. Wang et al [9] considers permanent or static immunization where a node once immunized is permanently protected. In reality, immunization must be taken as temporary due to multiple worm outbreaks since a computer being recovered from a certain worm can be reinfected by other worms immediately. In other words, any computer could not be permanently immune to many Internet worms.

In order to defend against future worms, we need to understand the network characteristics of worm spreading. Clearly the effect of factors such as the rate and pattern of infection, the underlying network topology, and human countermeasures in the network must be well understood before the model of Internet worm propagation could be developed. Certain nodes of the Internet are well protected compared with the others. Moreover, at certain vital installations the rates at which infections are cured are higher compared with others. To model these real world phenomenons we have taken into account in our simulations variable infection rates and variable cure rates.

Also in this paper, with real Internet topology data, we find that there are two effective factors that influence worm propagation: temporary immunization time and network delays. We note that our simulation results can explain how fast a virulent worm can spread and suggest effective mechanisms to monitor and defend against the propagation of worms. It also shows that we can find location(s) in the network that when quarantined would slow down the rage of spread.

The rest of the paper is organized as follows. Section 2 reviews the several related works. In Section 3, we give a brief review of the classical epidemic models and point out their limitations to model Internet worm propagation. In Section 4 and 5, we show the simulation results based on different network topologies. We conclude the paper with an outline of our future work in section 6.

## 2. Related Work

In epidemiology research, there exist several deterministic and stochastic models for virus spreading. About ten years ago, Kephart and White [3] presented the Epidemiological model (SIS) to understand and control the prevalence of viruses. This model is based on biological epidemiology and uses nonlinear differential equations to provide a qualitative understanding of virus spreading.

The Code Red worm incident of July 2001 has been investigated to model and analyze Internet worm propagation [1]. Based on the classical epidemic models, Zou et al [8] introduced a new general Internet worm model called *two-factor worm* model: one is the effect of human countermeasures against worm propagation; the other is the slower worm infection rate due to Internet congestion caused by Code Red worm.

Chen et al [10] present a model, referred to as the Analytical Active Worm Propagation (AAWP) model that characterizes the propagation of worms that employ random scanning. The AAWP model shows that the model can be applied to monitoring, detecting and defending against the spread of active worms in comparison of Weaver's simulation [11,12].

Wang et al [16] introduced an analytic model to capture the impact of underlying topology in computer viral propagation. They assume that an infection rate for each edge and a cure rate for each infected node are constant. In addition to the spread of a virus in real network, Wang and Wang [17] investigated the model extending the classical epidemic model by including two specific parameters: infection delay and user vigilance time. The infection delay is a period of time between the arrival of a virus on certain node and further infection from that node. The user vigilance time is the immune time. We examined several major characteristics of infection, including the variant rate and pattern of infection through the different network topologies and the rate of re-infection at individual nodes during an attack. We use a discrete time model and deterministic approximation to describe the spread of Internet worms.

## **3.** Worm Propagation Models

We introduce two classical deterministic epidemic models and an extension of one of models, which are the basis of our experimental design. We also point out their limitations when we try to use them to model Internet worm propagation. In classical epidemic model, it is defined that a host is called an *infectious* host at time t if it has been infected by virus before t. A host that is vulnerable to virus is called a *susceptible* host. We define that the temporary immunity is a temporary hold on a worm spreading, which means that many hosts will be susceptible or infected by new worm outbreaks at time t though they are already immune to old worm that came out before time t.

### 3.1. Classical Simple Epidemic Model

In classical simple epidemic model, each host stays in one of two states: susceptible or infectious. Each susceptible host becomes an infectious one at a certain rate. At the same time, infectious hosts are cured and become again susceptible at a different rate. This model system where having the infection and being cured does not confer immunity. This model is called the SIS model. Using the terms defined in Table 1, the differential equation for the SIS model is

$$\frac{dI(t)}{dt} = \beta I(t)[N - I(t)] - \delta I(t)$$
(1)  
$$\frac{dS(t)}{dt} = -\beta S(t)[N - S(t)] + \delta[N - S(t)]$$

The solution to the Equation 1 is

$$I(t) = \frac{I_0(\beta N - \delta)}{I_0\beta + (\beta N - \delta - I_0\beta)e^{-(\beta N - \delta)t}}$$
(2)

We conclude that, as  $t \to \infty$ ,

$$I_{\infty} = N - \rho \tag{3}$$

where  $\rho = \delta/\beta$  and  $I_0$  is the initial number of infectious hosts. Therefore, not absolutely all the population gets infected. This shows that each infectious host infects on average  $\beta$  others per unit time. However, the probability that a host becomes infected is not the same for every host

### Table 1. Notations of worm epidemic models

Notation	Definition
Ν	Size of total vulnerable population
S(t)	Number of susceptible hosts at time <i>t</i>
I(t)	Number of infectious hosts at time t
R(t)	Number of removed infectious hosts at time t
β	Infection rate
δ	Curing rate on an infectious host
λ	Removal rate on an infectious host
μ	Re-susceptible rate on a removed host
ρ	Epidemic threshold

because it is a function of their connectivity and the infection characteristics with a certain cure rate. We note that the probabilities per unit time of infection and of cure are independent. Once a host is cured, it is immediately capable of being reinfected. The SIS model does not take into account the possibility of individual's removal due to death or immunization which would lead to the so-called *Susceptible-Infectious-Removed* (SIR) model [6, 9]. It also does not model secondary effects such as reduced infection rate due to network congestion when many hosts are infected [8].

### 3.2 Kermack-Mckendrick Model

In epidemiology modeling, Kermack-Mckendrick model considers the removal process of infectious hosts [6]. This model is called the classical SIR epidemic model. Each host is assumed to be in one of three states: *Susceptible* (S), *Infectious* (I), and *Removed* (R).

$$\frac{dI(t)}{dt} = \beta S(t)I(t) - \gamma I(t)$$

$$\frac{dS(t)}{dt} = -\beta S(t)I(t) \qquad (4)$$

$$\frac{dR(t)}{dt} = \gamma I(t)$$

where  $\beta$  is the infection rate;  $\gamma$  is the rate of removal; N is the size of vulnerable population.

The Kermack-Mckendrick model improves the SIS epidemic model by considering that some infectious hosts are immune, are placed in isolation, or have died. However, this model is still not suitable for capturing the effect of multiple worm propagation simultaneously. First, in the Internet, many new viruses and worms come out every day though most of them disappear due to human countermeasures. In other words, many hosts will be susceptible or infected by new virus outbreaks at time t though they are already immune to recovered old virus that came out before time t. The link delays required for the infection to travel to the hosts are captured in the aggregate value called infection rate. While such gross estimates are correct for long lasting worms, it does capture neither the short lived ones nor the vulnerability of nodes which are reachable quickly.

### 3.3 An Extension for the SIR Model

We assume that a more general case, allowing for loss of immunity that causes recovered hosts to become susceptible again. In other words, a portion of the removed hosts a time t, R(t), due to loss of immunization join the susceptible population at time  $t + \tau$ ,  $S(t+\tau)$ . Therefore a portion of population dynamically changes from susceptible to infectious, to removed and back to susceptible.

Model that describes such an epidemical cycle is referred to as SIRS model.

Our model is a generalization presented in [7], allowing hosts recovering from the infective to go into a temporarily immune state rather than directly back into the susceptible state. Let  $\mu$  be the rate at which removals loose the immunization and becomes susceptible. Using the same notation as the SIR model we obtain the following deterministic SIRS model:

$$\frac{dS(t)}{dt} = -\beta I(t)S(t) + \mu R(t)$$

$$\frac{dI(t)}{dt} = \beta I(t)S(t) - \lambda I(t)$$

$$\frac{dR(t)}{dt} = \lambda I(t) - \mu R(t)$$
(5)

Also, we have S(t) + I(t) + R(t) = N,  $\forall t \ge 0$ . We can supply the same initial conditions as with the SIR model and numerically solve the SIRS model. Let  $\rho = \lambda/\beta$  be the epidemic threshold and  $I_0$  and  $S_0$  are the initial fraction of infectious hosts and of susceptible hosts, respectively. For the epidemic to occur, we must have:

$$\frac{dI}{dt}\Big|_{t=0} > 0 \to \beta S_0 I_0 - \lambda I_0 > 0 \to S_0 > \frac{\lambda}{\beta}$$
(6)

Clearly  $S_0$  must satisfy this condition for the epidemic to occur. The Eq. 6 indicates that no epidemic occurs if the initial number of susceptible hosts is smaller than the epidemic threshold,  $S_0 < \rho$ . This important result of the threshold effect is the same as what was already discovered by Kermack and McKendrick [6]; the population must be "large enough" for a disease to become epidemic.

Figure 1 compares the number of infectious, susceptible, and removed hosts as a function of time as obtained from Eq. 5. We attempt to solve this model using the numerical capabilities of MAPLE (mathematics software) without finding an explicit function-formula for the number of susceptible, infectious and removed hosts.

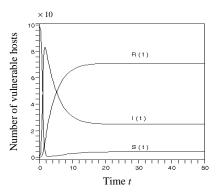


Figure 1. SIRS epidemic model; it shows the number of infectious, susceptible, and removed hosts as a function of time

The graph contains 100 hosts and the infection, removal and re-susceptible rates are  $\beta = 0.5$  and  $\lambda = 0.2$ ,  $\mu = 0.07$  respectively. It shows that the number of infectious hosts is initially exponentially increased up to about 80% of total population and then decreasing the growth of infection population. It is also observed that the infection growth will reach to a stable equilibrium after an amount of time passes.

While there is a vast literature covering models in which the "temporary immunity" step is not considered (i.e., SIS models and SIR models), comparatively little work has been done to understand how the nature of the  $R \rightarrow S$ transition affects the dynamics of an epidemic of Internet worms. With regard to the loss of immunity we consider two different types of worm behaviors, depending on parameters: (i) periodic epidemic outbreaks and (ii) one or more extended outbreaks followed by extinction of the epidemic due to stopping spreading of old worms. We note that instead of acquiring infinite immunity to a specific epidemic, infected hosts in this model spend a constant number of time steps in a generalized immune state before returned to the susceptible population. We have to investigate the SIRS model with immunity lasting nonconstant time step since hosts can be significantly delayed in the removed state by mechanisms such as a large constant period of temporary immunity.

# 4. Simulation and Analysis

In this section we present measurements of worm infections in two different network topologies with random rates at which an infectious node attempts to infect its neighboring nodes and random rates at which it protects itself or remove viruses itself. These experiments provide insight into the characteristics of infection propagation on computer networks and they also serve as the basis for future research work on quarantine of virulent Internet worms.

# 4.1. Random Transit-Stub Model without Topology Constraint

Our experiments have been conducted using a simulation environment that is capable of simulating hundreds of computing nodes with random network topology and any viral epidemic model. The network topology that is used in this simulation is constructed by Transit-Stub model that produces hierarchical graphs in a different way by consisting of interconnected transit and stub domains [14]. In this experiment we do not consider the topology constraint such as infection delay time when infective messages are able to reach a susceptible node. Instead, the infection process was simulated by varying the

connectivity of topology, the number of nodes, and the rate of infection  $\beta$  and cure  $\delta$ .

# 4.2. System Model

We consider a network with 100 nodes and two simulation scenarios. The first one is cured and infection case (CI strategy), the same as the one used in the classical simple SIS model, in which an infectious node determines whether it can be cured of infection or not before infecting any of susceptible nodes connected to it. The second one is infection and cured case (IC strategy) where an infectious node determines whether it can be cured or not after infecting any of susceptible neighboring nodes. We also analyzed the worm epidemic model with two different infection and cure rates: one is constant infection/cure rate at which an infectious node is equally likely to infect any of other susceptible nodes and to be cured of infection. The other one is variant infection/cure rate at which certain infectious nodes are likely to infect more susceptible nodes than other infectious nodes do. In addition, the rate of infection is associated with each of edges. Similarly, the rate of cure of infection is related to each node.

A few assumptions and simplifications were made to ensure feasibility of our experiment. First, a single initially infected node is randomly selected to release worm in each trial and we performed 500 simulation runs using same parameters. Second, a desired random graph has average degree of 5 on each node. In addition, relevant data is recorded per unit time and simulation stops when some desired state is reached, such as all nodes are infected or expiration of simulation.

# 4.3. Initial Results

Figure 2 shows the total number of infectious nodes averaged across the 500 runs of the two different types of simulation models. Note that the number of infectious nodes quickly reaches to almost 80% of the total population, and the infection growth slows down after that. This result is consistent with the results in simulation presented by Kephart [2]. Also considerable fraction of the nodes in a transit-stub network remains uninfected for long periods of time due to their connectivity. Comparing the two different epidemic strategies between constant infection/cure rate and non-constant infection/cure rate, we note that there is a slightly difference between these two strategies as shown in Figure 2 (a) (b). Also more rapid propagations were observed when different infection and cure rates are assigned to different nodes. Figure 2 (c) shows the total number of infectious nodes averaged over time with various temporary immunization times. As defined in section 3, we account for the temporary immunization time as the time period to protect the same infective messages from infectious nodes until new worm

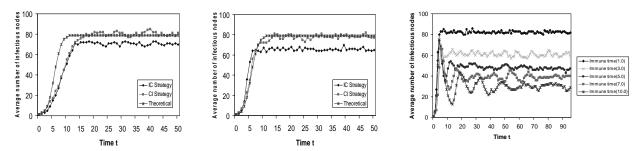


Figure 2. Comparison of the number of infectious nodes as a unit time in two different epidemic strategies; N = 100, (a) with constant rates,  $\beta$  = 0.5, and  $\delta$  = 0.2 (b) with non-constant rates (c) with constant rate and various temporary immunization times

comes out during a worm infection. We note that the difference of time between two worm outbreaks is referred to as the temporary immunization time. For instance, if an infectious node is cured of infection and takes 10 unit temporary immunization times at time t, it could be susceptible or reinfected by other worms at time t + 10. Results of these experiments show that for a given epidemic model the longer temporary immunization time, the wider will be the variation in infection growth. It is also observed that the infection growth of any type of propagation will reach to a stable equilibrium after an amount of time passes, which is consistent with the numerical solution obtained from SIRS model.

# **5. Internet Worm Propagation with Topology Constraint**

We extend our simulation methodology to include a realistic network model and evaluate the impact of topological constraints. After infecting a susceptible node, a worm attempts to infect other susceptible nodes with infection delay time which is the time to find its target nodes; it may attempt to only infect a small number of other susceptible nodes corresponding to network topological criteria, such as connectivity of network. In addition, we focus on the behavior of Internet worm propagation in response to multiple worm outbreaks. We model the impact of multiple active worms by specifying the temporary immunization time under which an infected node could be immune to the same type of worm after being cured.

# 5.1. Network Model

In this section, we describe the experimental network model of Internet worm infection using real Internet data set (round trip time (RTT) data) called topology constraint. In this study, we obtain network topology data (*e.g.* RTT data and traceroute) from the NLANR Active Measurement Program (AMP) [18]. Our network model consists entirely of 130 active measurement nodes

provided by AMP. Each node is connected to the global network shown in Abilene network topology [19].

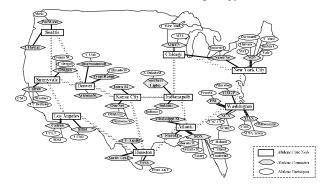


Figure 3 The Abilene Network Topology including Abilene core nodes, connectors and some of participants [19]

## 5.2. Simulation and Results

For our simulation, we set the discrete interval time into one millisecond (ms), the maximum simulation time for trial to 127 ms corresponding to the maximum RTT value observed from AMP. For Internet worm epidemic model we assume that the infection rate  $\beta$  and the cure rate  $\delta$  are the same as what are applied in the classical epidemic model. Figure 4 shows the result for the comparison of the total 60% infection times as the starting node in Internet

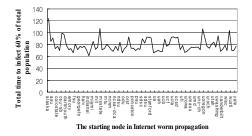


Figure 4. Comparison between the total times to infect 60% of total population vs. the starting node in a worm spreading with constant cure rate

worm propagation, obtained from 500 runs of the simulation for our network topology model. We note that some variation exists in the time to infect a large portion of the network. The fastest total time to infect 60% of total participant hosts is 61ms when Indiana University is infected first and the longest was 125ms when University of Alaska was infected first. As shown in Figure 3, Indiana University is considered as the center of Abilene backbone network as well as one of Abilene connectors connecting directly to the Abilene network while University of Alaska is just participant located away from the central point of the Abilene network.

We also measured the total number of re-infections that each participant host experienced during a worm infection in order to validate whether the structure of network topology has great influence on infection propagation. Figure 5 shows the result for the number of re-infections for each of the 130 participant belonging to Abilene network. We see that some number of the hosts is reinfected much more than others. For example, Indiana University is attacked 86 times on average, while Wayne state university is attacked 41 times only.

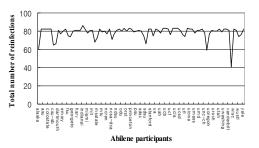


Figure 5. Counting the total number of reinfections at each participant host

It may be pointed out that for a given topology we might slow down the growth of Internet worm infection if we find critical locations where some nodes are more prone to being attacked more than others. Moore et al [15] investigated the containment system using address blacklisting and content filtering to minimize worm propagation in the Internet. The simulation system we have performed could identify addresses to be blacklisted.

## 6. Conclusion

In this paper we have presented measurements of worm infections in two different network topologies with constant or non-constant infection and cure rates. We extended our simulation methodology to include a real Internet network model and evaluated the impact of topology constraints. We are also working on the development of effective quarantine techniques using the knowledge of worm propagation.

## References

- D. Moore, C. Shanning, and K. Claffy, "Code-Red: a case study on the spread and victims of an Internet worm", *Proceedings of the 2nd Internet Measurement Workshop*, November 2002, pp. 273–284.
- [2] J. O. Kephart and S. R. White, "Directed-graph epidemiological models of computer viruses", *Proceedings of the IEEE Symposium on Security and Privacy*, 1991, pp. 343–361.
- [3] J. O. Kephart and S. R. White. "Measuring and Modeling Computer Virus Prevalence", *Proceedings of the IEEE* Symposium on Security and Privacy, 1993.
- [4] E. H. Spafford. "The internet worm incident", In ESEC'89 2<sup>nd</sup> European Software Engineering Conference, Coventry, United Kingdom, 1989.
- [5] F. Cohen, "Computer Viruses: Theory and Practice", Computers & Security, vol. 6, pp. 22–35, Feb. 1987.
- [6] J. C. Frauenthal. *Mathematical Modeling in Epidemiology*, Springer-Verlag, New York, 1980.
- [7] L. Edelstein-Keshet, *Mathematical Models in Biology*, Random House, New York, 1988.
- [8] C. C. Zou, W. Gong, and D. Towsley, "Code Red Worm Propagation Modeling and Analysis", *Proceedings of the* 9th ACM Conference on Computer and Communications Security, November 2002, pp. 138–147.
- [9] C. Wang, J. C. Knight, and M. C. Elder, "On Computer Viral Infection and the Effect of Immunization", *Proceedings of the 16th Annual Computer Security Applications Conference*, 2000, pp. 246–256.
- [10] Z. Chen and L. Gao and K. Kwiat, "Modeling the Spread of Active Worms", *IEEE INFOCOM*, 2003.
- [11] N. Weaver. "Warhol Worms: The Potential for Very Fast Internet Plagues", http://www.cs.berkeley.edu/~nweaver/ warhol.html.
- [12] S. Staniford, V. Paxson, and N. Weaver, "How to Own the Internet in Your Spare Time", *Proceedings of the* USENIX Security Symposium, August 2002, pp. 149–167.
- [13] eEye Digital Security. Press. "New Critical Flaw in Microsoft Windows® RPC", <u>http://www.eeye.com/html/Press/PR20030910.html</u>.
- [14] Ellen W. Zegura, Ken Calvert and S. Bhattacharjee. "How to Model an Internetwork", *Proceedings of IEEE INFOCOM '96*, San Francisco, CA.
- [15] D. Moore, C. Shannon, G. Voelker, and S. Savage, "Internet quarantine: Requirements for containing selfpropagating code", *Proceedings of 22nd Annual Joint Conference of IEEE Computer and Communication* societies (INFOCOM 2003), April 2003.
- [16] Yang Wang, Deepayan Chakrabati, Chenxi Wang, and Christos Faloutsos, "Epidemic spreading in real networks: an Eigenvalue viewpoint", *Proceedings of 22nd International Symposium on Reliable Distributed Systems*, October, 2003.
- [17] Yang Wang, Chenxi Wang, "Modeling the Effects of Timing Parameters on Virus Propagation", *Proceedings of* the 2003 ACM workshop on Rapid Malcode (WORM 03), October, 2003
- [18] NLANR Active Measurement Program (AMP), http://watt. nlanr.net/
- [19] Abilene network, http://abilene.internet2.edu/