Effectiveness of Quarantine in Worm Epidemics

Thomas M. Chen and Nasir Jamil
Department of Electrical Engineering
Southern Methodist University
Dallas, Texas 75275
Email: tchen@engr.smu.edu, nasir@mail.smu.edu

***Abstract*— Quarantine is a natural concept borrowed from human disease control to slow down worm outbreaks. We study the effectiveness of partial quarantine for simple epidemics (without removals) and find that the optimal quarantine strategy is not as simple as expected. The strategy depends on which networks are most important to protect. We also investigate the effectiveness of quarantine for general epidemics (with removals) and derive the critical threshold for networks to have herd immunity. We show that, given a limited capability to quarantine a given number of networks, the optimal quarantine strategy is to isolate the networks small enough to have herd immunity, and then divide the remaining networks as evenly as possible.**

I. Introduction

Quarantine is a natural and widely practiced method of human disease control. Since many diseases are transmitted from infectious to susceptible individuals through social contact, an epidemic can be curtailed by isolating the infectious subpopulation. In practice, it may not not be possible to isolate all the infectious individuals. Thus, the basic goal of quarantine modeling is to to gain an understanding of the effectiveness achieved by quarantining some subset of infectious individuals [4].

The same concepts apply to quarantine of network worms [16]. Worms are automated programs that take advantage of network connectivity to spread from infected hosts to vulnerable hosts. In theory, worm traffic can be blocked and filtered by firewalls, intrusion prevention systems, and routers with access control. Quarantining a subnetwork will prevent infected hosts within the subnetwork from infecting vulnerable hosts in other networks as well as prevent external hosts from infecting internal vulnerable hosts.

Quarantine is only one defense among many and insufficient by itself. Other defenses are needed to fortify hosts by removing vulnerabilities and cure infected hosts through software patching and antivirus software. Quarantine serves to slow down an outbreak to buy time for vulnerable hosts to be fortified and infected hosts to be disinfected. In major worm outbreaks in the past, the patching and cleanup activities were employed when the worm epidemic had already progressed to an advanced stage. Quarantine may slow down an epidemic sufficiently to apply these defensive activities in the early stages of the epidemic, thus minimizing the ultimate damage.

An alternative to quarantine is rate throttling, advocated by Williamson and others [21]. Williamson postulated that normal applications exhibit a stable contact rate (found to be less than 5 connections per sec.) to a limited number of external hosts (servers). His virus throttling approach keeps an active set of addresses for each host. Outbound connections to these addresses are allowed, but other outbound connections are delayed by putting them in a queue that is serviced at a rate of 1-2 per second. Thus connections to frequently contacted addresses are allowed but connections to random new addresses are delayed. Wong et al. examined the effectiveness of rate throttling performed at routers as well as hosts [22].

In practice, the effectiveness of quarantining will depend on the time to detect a new worm outbreak and activate quarantine. A worm can spread without constraint among the vulnerable subpopulation before it is detected. It is critical to minimize the time to detect a new worm outbreak. An extensive amount of literature on intrusion detection addresses the issue of automated worm detection [2], [5], [6], [8], [9], [14], [17]–[19].

The effectiveness will also depend on the extent of deployment of filtering equipment within the network. Quarantining will most likely be implemented at firewalls at network edges, but not universally. This paper investigates the effectiveness of partial quarantine and the question of optimal quarantine strategies. Section 2 studies the impact of quarantine for simple epidemics (without removals). Section 3 examines quarantine for general epidemics (with removals). A central concept for general epidemics is herd immunity. We derive a critical threshold for quarantined networks to have herd immunity, and examine the impact of herd immunity on the optimal quarantine strategy.

II. Quarantine for Simple Epidemics

Research on worm quarantine strategies is still at an early stage. Moore et al. examined a topology map of autonomous systems in the Internet and compared deployment of content filtering at large ISPs versus customer autonomous systems [16]. It was concluded that a worm outbreak can be contained to a minority of hosts if the top 20 ISPs can block the worm traffic. In this paper, we seek more general results that are not specific to a particular Internet topology.

Zou et al. investigated a "soft" quarantining scheme where hosts suspected of being infected are quarantined temporarily for some length of time (but could be re-quarantined) [24]. This was evaluated by the simple and general epidemic models.

Zou et al. proposed a "firewall network system" consisting of internal firewalls for dividing an enterprise network into

isolated subnetworks [25]. The quarantining works with an active patching system that aggressively identifies and patches vulnerable hosts. The study focuses mostly on implementation and architectural issues.

Liljenstam et al. compared quarantine with active patching and "counter-worms" [13]. It was not clear why quarantine is compared with patching, since they are complementary and not opposing methods. It was concluded that quarantining needs to be deployed very widely and act very quickly in order to be effective. In this paper, we view quarantine as a complementary method to patching. Quarantine serves to buy time for systems to be patched and fortified.

A. Simple Epidemics

The "simple epidemic" or SI (susceptible $\rightarrow$ infective) model assumes a homogeneous vulnerable population [1], [7]. The population is considered to be a fixed number of N hosts during the timeframe of interest. The population is initially entirely susceptibles (i.e., vulnerable but not infected) except for a small number of infectives. Through contacts with infectives, susceptibles may become infective and then remain infective permanently.

Let $S(t)$ and $I(t)$ denote the number of susceptibles and infectives at time t, where $S(t)+I(t)=N$. By homogeneous mixing, each susceptible is assumed to make an average βN contacts per unit time but the probability of meeting a susceptible each time is S/N. The parameter β is the pairwise infection rate or infectious contact rate. Hence, the number of infectives increases at a rate of

$$\frac{d}{dt}I = (\beta N)(S/N)I = \beta IS = \beta I(N-I) \tag{1}$$

Given the initial condition $I(0)=I_0$, the solution is the logistic curve

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

According to (2), an outbreak will reach an infection level pN at time

$$T_p = \frac{\ln p(N-I_0) - \ln I_0(1-p)}{\beta N} \tag{3}$$

The SI model appears to be a good candidate for early stages of random scanning worm epidemics. These worms target pseudo-random IP addresses which seems to conform to the assumption of homogeneous mixing. Moore et al. showed that the logistic curve predicted by the SI model could fit the observed data for the growth of the Code Red worm [15]. Liljenstam et al. fit the SI model to the initial spread of the SQL Slammer worm [12]. Zou et al. agreed with the close fit for the early stages of the Code Red outbreak but pointed out a greater than predicted slowdown in the later stages [23]. The discrepancy in the later stages was attributed to the fact that the SI model did not account for network congestion and human countermeasures (such as patching, filtering and isolation).

B. Effectiveness of Partial Quarantine

The effect of quarantine is to divide the population into separate subpopulations which do not mix with each other. Liljenstam et al. modeled the Internet as an interconnected set of networks or autonomous systems [11]–[13]. Wagner et al. also chose to model worm propagation through an Internet structured as an interconnection of multiple subnetworks [20]. Worm quarantine prevents infections from spreading from one network to another. However, if a quarantined network is already infected, the epidemic will continue to spread within that network even after the quarantine. Even so, the spread of infection within the network is going to slow down significantly after quarantine, because there will be no contribution of infectious contacts from other networks.

For the moment, we assume that the population consists of m networks which are all equal size $N_1 = \cdots = N_m = N/m$ (this assumption will be relaxed later). As a practical matter, only a fraction P of networks will be able to be quarantined. The unquarantined networks spread infections between them without constraint, while quarantined networks have spreading only within each network. We used parameter values estimated by Liljenstam et al. for the SQL Slammer worm: $\beta = 5.6 \times 10^{-5}$, $N = 120,000$ [12]. For $m = 100$, Fig. 1 shows the epidemic rates as a function of P. As might be expected, larger values of P (more quarantined networks) cause the epidemic to slow down more. In each case, there is a very fast initial spread dependent on the number of unquarantined networks that spread infections between them, followed by a much slower spread contributed by the spread of infections within the quarantined networks.

Fig. 2 shows the time to saturate a certain fraction of the population as a function of P. The time to infect 95 percent of the population is slowed very substantially even for small P. However, it takes much more quarantining (larger P) to have a significant effect on the time to reach lower infection levels. For example, it takes more than 75 percent quarantining to

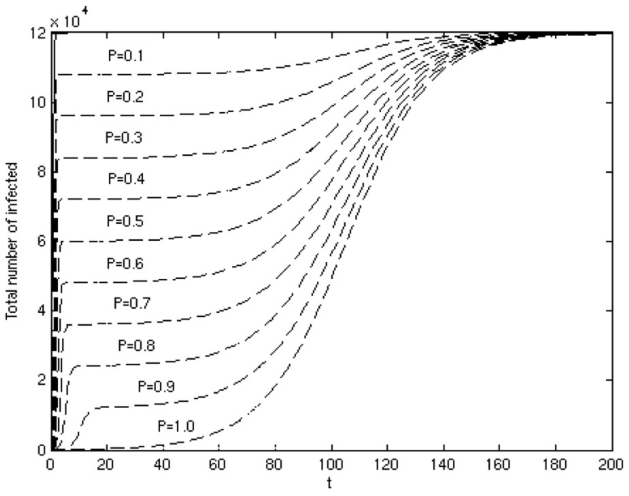

Fig. 1. Epidemic rates as a function of P

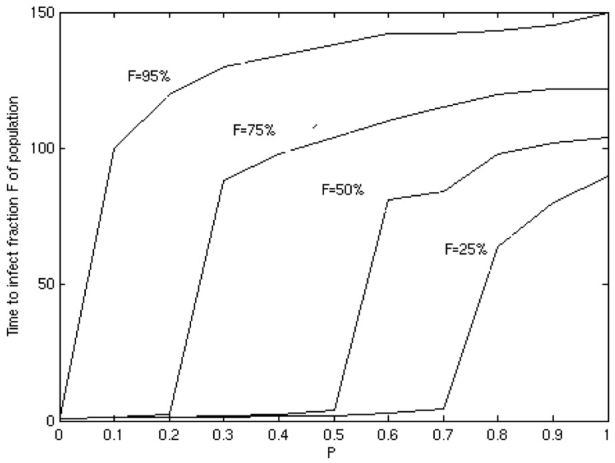

Fig. 2. Effect of P on time to saturate a certain fraction of the population

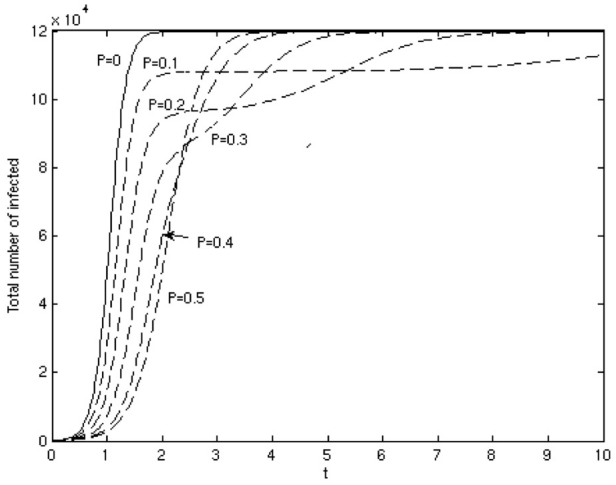

Fig. 3. Total number of infectives over time as a function of P

have an effect in slowing down the time to infect 25 percent of the population. The reason is that 75 percent quarantining leaves 25 percent of the population unquarantined, and this 25 percent subpopulation will saturate quickly.

C. Strategies for Quarantine

We now relax the assumption of equal size households and consider households of different sizes to investigate the question of optimal quarantining strategy. It will not be practical to quarantine every network because this would require worm-blocking firewalls at the edge of every network. Suppose only a given number of firewalls are capable of worm quarantine, which networks should be quarantined? Our intuition for a reasonable strategy is to quarantine the largest networks first, which would leave the smallest subpopulation unquarantined. However, the situation is not that simple.

We follow an inductive argument to examine this strategy. We first consider one network to quarantine, then the second network, and so on. The first quarantined network will separate the population into two subpopulations. Let us suppose the two subpopulations have sizes PN and $(1-P)N$. That is, the first network to quarantine represents a fraction P of the population. The epidemics in the two subpopulations grow independently.

Fig. 3 shows the total number of infected over time as a function of P. For each value of P, there is an initial fast epidemic growth due to the larger subpopulation, followed by a relatively slow growth due to the smaller subpopulation. The figure shows the net growth of infection in both the networks. The initial epidemic growth is slowest when $P = 0.5$ or the subpopulations are the same size. However, the later epidemic growth is slowest when the subpopulations are very different in size. A small value of P is more advantageous because the small network of size PN is very slow to saturate, which keeps the total population from saturating completely. When there is a large subpopulation and a small subpopulation, the large subpopulation does saturate relatively quickly but the small subpopulation is much slower to saturate. This slows down the saturation in the later stages of the epidemic.

Fig. 4 shows the effect of P on reaching the 50 percent and 95 percent infection levels. Small values of P means there is a large subpopulation and a small subpopulation. The large subpopulation saturates quickly, while the small subpopulation saturates very slowly. Therefore it takes a long time for the epidemic to reach an overall 95 percent infection level. For larger values of P, the two subpopulations are more equal in size. Both will then saturate at moderate rates. The time to reach the 95 percent infection level becomes shorter.

Fig. 5 shows the time for a single subpopulation of size N to saturate to a 95 percent level. As the subpopulation size decreases, the time to saturate increases exponentially. This points out that whenever possible, it is best to quarantine a population into the smallest possible subpopulations.

The trade-offs involved in quarantine strategies are shown in Fig. 6. If the smallest networks are quarantined first, they

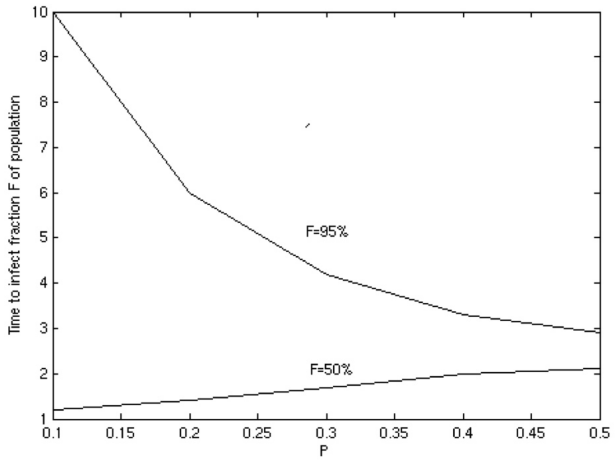

Fig. 4. Effect of P on time to achieve a certain infection level

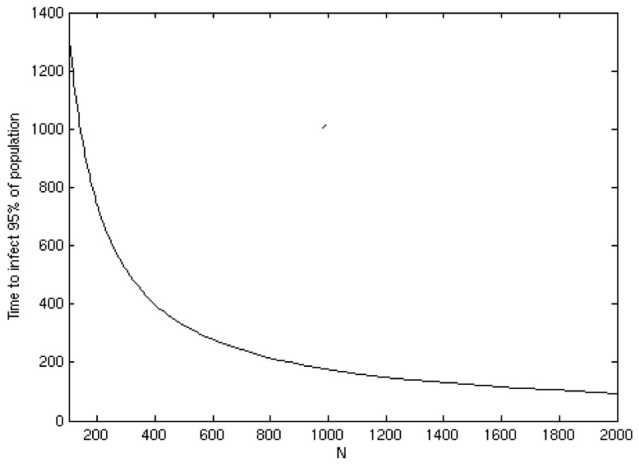

Fig. 5. Effect of N on time to achieve 95 percent infection level

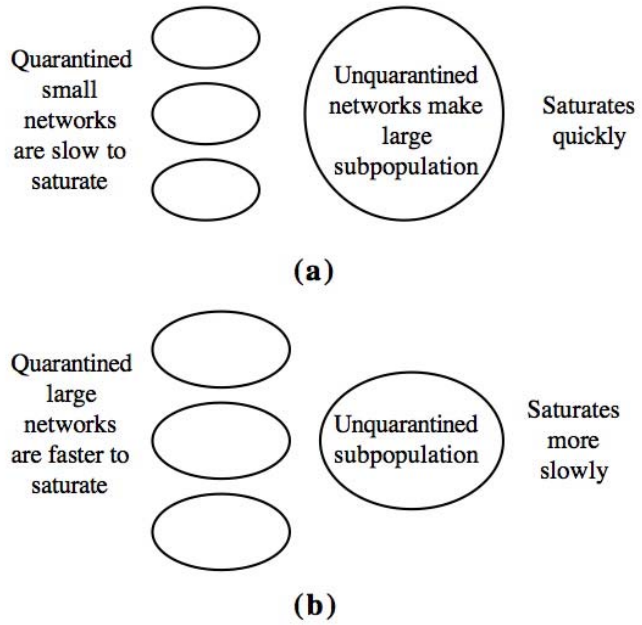

Fig. 6. Illustration of trade-offs involved in quarantining strategies

will be well protected and slow to saturate. At the same time, however, it leaves a relatively large subpopulation which will saturate quickly. Fig. 6(a) is the best strategy if the protection of the small quarantined networks is of paramount importance. The alternative is to quarantine the largest networks first. The large quarantined networks will saturate more quickly. At the same time, a relatively smaller unquarantined subpopulation will saturate at a moderate rate. Fig. 6(b) is the best strategy if protection of the unquarantined subpopulation is as important as protection of the quarantined networks.

III. Quarantine for General Epidemics

The SI model is good for fast epidemics that can spread without initial constraints. In other cases, there will be countermeasures such as software patching and antivirus disinfecting during a worm outbreak. Zou et al. noted that the simple epidemic model is a close fit for the early stages of the Code Red outbreak but does not account for network congestion or human countermeasures in the later stages [23].

A. General Epidemics

The general epidemic or SIR (susceptible → infective → removed) model due to Kermack and McKendrick [10] has been used for worm epidemics by several researchers [11], [12], [23], [24]. For a closed population, the number of susceptibles $S(t)$, infectives $I(t)$, and removed $R(t)$ are governed by the system of differential equations

$$\frac{d}{dt}S = -\beta SI \quad (4)$$

$$\frac{d}{dt}I = \beta SI - \gamma I \quad (5)$$

$$\frac{d}{dt}R = \gamma I \quad (6)$$

The SIR model is similar to the SI model except for the additional transition of infectives to removed state. The removal rate γ reflects human countermeasures to disinfect, patch, or disconnect infected hosts.

Let $\{N_1, \ldots, N_m\}$ denote the sizes of m networks, and $N = N_1 + \cdots + N_m$ is the total population. Network j has $p_j = N_j/N$ fraction of the total population. The number of susceptibles, infectives, and removed in network j are $S_j(t)$, $I_j(t)$, and $R_j(t)$, respectively.

We consider that a worm epidemic had a time T already to spread without constraint before a quarantine begins. This delay is due to the time needed to detect a new outbreak. During this initial spread, the epidemic is spreading homogeneously. We assume that the epidemic started at time $t = -T$ and that the quarantine is activated at time $t = 0$. The initial unconstrained spread has the effect of establishing an initial number of infectives $I(0) = I_0$. Assuming that the epidemic originated at a single host, the total number of infectives at time $t = 0$ will be

$$I_0 = \frac{N}{1 + (N-1)e^{-\beta NT}} \quad (7)$$

Since the initial spread was homogeneous, the initial conditions for network j will be

$$I_j(0) = p_j I_0,\ S_j(0) = N_j - I_j(0),\ R_j(0) = 0 \quad (8)$$

B. Herd Immunity

An important result for SIR epidemics is the critical threshold for herd immunity. Notice that

$$\frac{d}{dt}I = (\beta S - \gamma)I \quad (9)$$

and hence, $I(t)$ will always decrease if $S(0) < \gamma/\beta$. The critical level γ/β is the initial number of susceptibles that would be sufficient to mix with the infectives for additional spreading. Below the threshold, $I(t)$ will decrease to zero. Above the threshold, the number of infectives will increase to

a maximum and then eventually approach an endemic level. The maximum level of infectives will be [3]

$$I_{max} = S(0) + I(0) - \frac{\gamma}{\beta}\ln S(0) - \frac{\gamma}{\beta} + \frac{\gamma}{\beta}\ln\frac{\gamma}{\beta} \qquad (10)$$

C. Strategies for Quarantine

Consider that a population consists of m networks, and only a given fraction of them can be quarantined. That is, we are able to selectively deploy a given number of worm-blocking firewalls at these networks. Each quarantined network will contain an SIR epidemic. If the initial number of susceptibles in a quarantined network is sufficiently small, that network will have herd immunity meaning the epidemic will always decrease and eventually disappear. The epidemic in network j will be governed by

$$\frac{d}{dt}I_j = (\beta S_j - \gamma)I_j \qquad (11)$$

and herd immunity is attained if $S_j(0) < \gamma/\beta$. After substitutions and rearrangement, the *condition for herd immunity* in a quarantined network is

$$N_j < \frac{\gamma}{\beta}\left(\frac{N}{N - I_0}\right) \qquad (12)$$

Therefore, if we can determine the epidemic rates β and γ and the extent of the epidemic I_0 at the start of quarantine, we can identify the networks smaller than critical size which have herd immunity. A reasonable strategy is to select these networks for quarantine because the epidemics in these networks will steadily dwindle without increase.

If we are able to quarantine some of the remaining networks above critical size, which should be selected? Again, we take an inductive approach and first ask how to quarantine a population into two subpopulations to minimize the total epidemic. We consider a total population of size N. Suppose we can quarantine this population into two subpopulations of size $N_1 = PN$ and $N_2 = (1-P)N$. The SIR epidemics in these two subpopulations will reach maximum levels

$$I_{1,max} = N_1 - \frac{\gamma}{\beta}\ln S_1(0) - \frac{\gamma}{\beta} + \frac{\gamma}{\beta}\ln\frac{\gamma}{\beta} \qquad (13)$$

$$I_{2,max} = N_2 - \frac{\gamma}{\beta}\ln S_2(0) - \frac{\gamma}{\beta} + \frac{\gamma}{\beta}\ln\frac{\gamma}{\beta} \qquad (14)$$

Putting everything in terms of P, this can be rewritten as

$$I_{1,max} = PN - \frac{\gamma}{\beta}\ln(PN - PI_0) - \frac{\gamma}{\beta} + \frac{\gamma}{\beta}\ln\frac{\gamma}{\beta} \qquad (15)$$

$$I_{2,max} = (1-P)N - \frac{\gamma}{\beta}\ln((1-P)N \qquad (16)$$

$$-(1-P)I_0) - \frac{\gamma}{\beta} + \frac{\gamma}{\beta}\ln\frac{\gamma}{\beta} \qquad (17)$$

If we minimize the total maximum $I_{max} = I_{1,max} + I_{2,max}$ with respect to P, we find that the optimal is $P = 1/2$. In other words, the subpopulation should be divided into equal subpopulations to minimize the total epidemic.

These results lead to an optimal quarantine strategy for quarantining general epidemics:

1) First select the networks below the critical size for herd immunity to quarantine
2) If more networks can be quarantined, selectively deploy firewalls to divide the remaining subpopulation as evenly as possible

IV. Conclusions

In this paper, we have examined the effectiveness of partial quarantines for simple epidemics and proposed quarantine strategies for simple and general epidemics. We have found that the size of quarantined networks has a great effect on the time for an outbreak to saturate the population. As the size decreases, the time to saturate increases exponentially. As a consequence, it is best to quarantine a population into the smallest possible subpopulations.

The optimal quarantine strategy for simple epidemics depends on which subpopulations are most important to protect. Quarantining the largest networks first leaves a relatively small unquarantined subpopulation. This is the best strategy when the unquarantined subpopulation is as important to protect as the quarantined networks. Quarantining the smallest networks is the best strategy if protection of the small quarantined networks is most important.

For general epidemics, we derived a critical threshold for quarantined networks to have herd immunity. It is advantageous to quarantine the networks smaller than the critical threshold because epidemics in these networks will dwindle without increase. We have shown that, if additional quarantining is possible, the remaining unquarantined population should be divided as evenly as possible, in order to minimize the total epidemic.

References

[1] N. Bailey, *The Mathematical Theory of Infectious Diseases and its Applications*, 2nd ed. NY: Oxford U. Press, 1975.

[2] V. Berk, G. Bakos, and R. Morris, "Designing a framework for active worm detection on global networks," in *First IEEE Int. Workshop on Info. Assurance (IWIAS 2003*, March 24, 2003, pp. 13-23.

[3] C. Castillo-Chavez, et al., eds., *Mathematical Approaches for Emerging and Reemerging Infectious Diseases: an Introduction*. NY: Springer-Verlag, 2002.

[4] C. Castillo-Chavez, C. Castillo-Garsow, and A-A. Yakubu, "Mathematical models of isolation and quarantine," *J. of Am. Medical Assoc.*, vol. 290, pp. 2876-2877, Dec. 3, 2003.

[5] S. Chen and S. Ranka, "An Internet-worm early warning system," in *IEEE Globecom 2004*, Dallas, TX, Nov. 29 - Dec. 3, 2004, pp. 2261-2265.

[6] X. Chen and J. Heidemann, "Detecting early worm propagation through packet matching," Technical Report ISI-TR-2004-585, USC/ISSI 2004.

[7] D. Daley and J. Gani, *Epidemic Modeling: An Introduction*, Cambridge, UK: Cambridge U. Press, 1999.

[8] G. Ganger, G. Economou, and S. Bielski, "Self-securing network interfaces," technical report CMU-CS-02-144, Carnegie Mellon U., Aug. 2002.

[9] G. Gu, et al., "Worm detection, early warning and response based on local victim information," in *20th Annual Comp. Sec. Applic. Conf.*, Dec. 6-10, 2004, pp. 136-145.

[10] W. Kermack and A. McKendrick, "A contribution to the mathematical theory of epidemics," *Proc. Roy. Soc. Lond. A*, vol. 115, pp. 700-721, 1927.

[11] M. Liljenstam, et al., "A mixed abstraction level simulation model of large-scale Internet worm infestations," in *10th IEEE/ACM Symp. on Modeling, Analysis, and Simulation of Comp. Telecom. Sys. (MASCOTS 2002)*, Fort Worth, TX, Oct. 11-16, 2002, pp. 109-116.

[12] M. Liljenstam, et al., "Simulating realistic network worm traffic for worm warning system design and testing," in *2003 ACM Workshop on Rapid Malcode (WORM)*, Wash. DC, Oct. 2003, pp. 24-33.
[13] M. Liljenstam and D. Nicol, "Comparing passive and active worm defenses," in *1st Int. Conf. on Quantitative Eval. of Sys. (QEST 2004)*, Enschede, Netherlands, Sept. 27-30, 2004, pp. 18-27.
[14] M. Martin, J.-M. Robert, and P. van Oorschot, "A monitoring system for detecting repeated packets with applications to computer worms," technical report TR-04-02, Carleton U., 2004.
[15] D. Moore, C. Shannon, and J. Brown, "Code-Red: a case study on the spread and victims of an Internet worm," in *2nd ACM SIGCOMM Workshop on Internet Measurement (IMW)*, Marseille, France, Nov. 6-8, 2002, pp. 273-284.
[16] D. Moore, et al., "Internet quarantine: requirements for containing self-propagating code," in *IEEE Infocom 2003*, San Francisco, CA, 2003, pp. 1901-1910.
[17] P. Porras, et al., "A hybrid quarantine defense," in *ACM Workshop on Rapid Malcode (WORM 2004)*, Wash. DC, 2004, pp. 73-82.
[18] S. Schechter, J. Jung, and A. Berger, "Fast detection of scanning worm infections," in *7th Int. Symp. on Recent Adv. in Intrusion Detection (RAID)*, Sophia Antipolis, France, Sept. 15-17, 2004.
[19] S. Singh, et al., "The EarlyBird system for real-time detection of unknown worms," technical report CS2003-0761, UCSD, 2003.
[20] A. Wagner, et al., "Experiences with worm propagation simulations," in *ACM Workshop on Rapid Malcode (WORM 2003)*, Wash. DC, Oct. 27, 2003, pp. 34-41.
[21] M. Williamson, "Throttling viruses: restricting propagation to defeat malicious mobile code," in *18th Annual Comp. Sec. Appl. Conf.*, Las Vegas, NV, Dec. 9-13, 2002.
[22] C. Wong, et al., "Dynamic quarantine of Internet worms," in *Int. Conf. on Dependable Sys. and Networks (DSN-2004)*, Florence, Italy, June 28 - July 1, 2004, pp. 62-71.
[23] C. Zou, W. Gong, and D. Towsley, "Code Red worm propagation modeling and analysis," in *9th ACM Conf. on Computer and Commun. Sec. (CCS'02)*, Wash. DC, Nov. 18-22, 2002, pp. 138-147.
[24] C. Zou, W. Gong, and D. Towsley, "Worm propagation modeling and analysis under dynamic quarantine defense," in *ACM Workshop on Rapid Malcode (WORM 2003)*, Wash. DC, Oct. 27, 2003, pp. 51-60.
[25] C. Zou, D. Towsley, and W. Gong, "A firewall network system for worm defense in enterprise networks," technical report TR-04-CSE-01, U. Mass. Amherst, Feb. 2004.