Deterministic and Stochastic Models for the Detection of Random Constant Scanning Worms

Kurt R. Rohloff
BBN Technologies
and
Tamer Başar
The University of Illinois

This paper discusses modeling and detection properties associated with the stochastic behavior of Random Constant Scanning (RCS) worms. Although these worms propagate by randomly scanning network addresses to find hosts that are susceptible to infection, traditional RCS worm models are fundamentally deterministic. A density-dependent Markov jump process model for RCS worms is presented and analyzed herein. Conditions are shown for when some stochastic properties of RCS worm propagation can be ignored and when deterministic RCS worm models can be used. A computationally simple hybrid deterministic/stochastic point-process model for locally observed scanning behavior due to the global propagation of an RCS scanning worm epidemics is presented. An optimal hypothesis-testing approach is presented to detect epidemics of these under idealized conditions based on the cumulative sums of log-likelihood ratios using the hybrid RCS worm model. This paper presents in a mathematically rigorous fashion why detection techniques that are only based on passively monitoring local IP addresses cannot quickly detect the global propagation of an RCS worm epidemic with a low false alarm rate, even under idealized conditions.

Categories and Subject Descriptors: G.3 [**Probability and Statistics**]: Stochastic processes; Time series analysis; D.4.6 [**Operating Systems**]: Security and protection—*Invasive software (e.g., viruses, worms, Trojan horses)*; I.6.8 [**Simulation and Modeling**]: Types of Simulation—*Discrete event*

General Terms: Security, Theory

Additional Key Words and Phrases: Worms, stochastic analysis, epidemic modeling, hypothesis testing

1. INTRODUCTION

A computer worm is a piece of malicious code that can spread automatically over a computer network without the need for human intervention. Due to this automatic propagation, worms can potentially spread on the Internet with staggering speed and cause damage on

This research was supported by the NSF grant CCR 00-85917 ITR.

Addresses:
Kurt R. Rohloff is with BBN Technologies, 10 Moulton St., Cambridge, MA 02138, USA, krohloff@bbn.com.
Tamer Başar is with the Coordinated Science Laboratory at The University of Illinois 1308 West Main St., Urbana, IL 61801, USA, tbasar@control.csl.uiuc.edu.

the order of billions of dollars [Moore et al. 2002; Moore et al. 2003]. A special type of a worm, called a Random Constant Scanning worm (RCS worm for short) propagates by continually scanning randomly selected Internet addresses in attempts to infect other hosts. Well-known examples of RCS worms include CodeRed1v2 and Slammer. When an Internet address has been selected for scanning by a host infected by an RCS worm, the infected host attempts to transmit infectious packets to a host at the selected address. If the targeted host is susceptible to the infection, then, upon receiving the infectious packets, that host also becomes infected and the scanning and infection process continues at both infected hosts.

Due to the branching nature of scanning worm epidemics, these infections have the potential to propagate very fast over the Internet. For instance, during the CodeRed1v2 Internet worm attack of 2001 over 359,000 computers were infected in under 14 hours [Moore et al. 2002]. During the more aggressive Slammer Internet worm attack of 2003 more than 90 percent of 75,000 vulnerable computers were infected in less than 10 minutes [Moore et al. 2003]. Although it might be feasible for a human to respond to a relatively slow RCS worm epidemic to protect a local network, human driven responses are infeasible for local network protections on a global scale, especially when vigilance is required twenty-four hours a day, seven days a week, fifty-two weeks a year. This realization motivates the need for a better understanding of RCS worms and the need for automated worm epidemic detection methods with response times and reliability significantly better than what could be provided by human network administrators.

From data collected during previous RCS worm attacks it has been found that the deterministic simple epidemic model [Kermack and McKendrick 1927] can capture aspects of the behavior of an RCS worm epidemic's propagation [Moore et al. 2002; Moore et al. 2003; Staniford et al. 2002]. This model was first introduced by Kermack and McKendrick [1927] for the modeling of biological epidemics when computational power for stochastic modeling was extremely limited. However, despite the use of a deterministic model, the underlying propagation behavior of RCS worms is fundamentally stochastic in nature.

It has been noted in the literature that due to the random nature in which an RCS worm spreads, there could be variability between the overall propagation rates of RCS worm epidemics for worms with similar propagation properties [Moore et al. 2003; Zou et al. 2003; Nicol 2006]. Of particular interest is the work by Nicol [2006]. Nicol [2006] presents a detailed model of CodeRed1v2 propagation which validates the density-dependent Markov jump process model previously introduced in [Rohloff and Başar 2005a] and discussed below extensively.

There has been some discussion of such stochastic effects in models for epidemics in the epidemiology literature (notably by Andersson and Britton [2000] and Mode and Sleeman [2000]), but to the best knowledge of the authors there has been no work analyzing the stochastic properties of RCS worm epidemics or justifying the use of the deterministic epidemic models in a mathematically rigorous fashion despite the inherently stochastic behavior of RCS worms.

This paper presents an idealized stochastic propagation model for RCS worms developed from first principles. This epidemic model is taken from the literature of epidemiology and public health [Andersson and Britton 2000]. The large-scale propagation behavior of an RCS worm predicted by this model is compared to the large-scale behavior predicted by the standard deterministic simple epidemic model. It is found that the major differences

between behavior predicted by the stochastic model and deterministic model is when the worm is in its earliest and latest stages of propagation when the observed differences between the behaviors of worms using these models is negligible. Conditions are shown for when some stochastic properties of RCS worm epidemic propagation can be safely ignored.

The deterministic simple epidemic model has been widely used in the literature as a basis for developing worm detection methods [Staniford 2003; Wong et al. 2004; Zou et al. 2003; Zou et al. 2004]. Based on our analyses of when some stochastic aspects of RCS worm epidemic propagation can be ignored, we present a hybrid deterministic/stochastic point process model for the observed scanning behavior on a local network due to the global propagation of an RCS scanning worm. Such a model has not been previously discussed in the literature.

Based on the hybrid deterministic/stochastic model for the observed scanning behavior on a local network due to an RCS worm epidemic, we present a cumulative-sum log-likelihood RCS worm epidemic detection method. We discuss the abilities of anomaly-based RCS worm epidemic detection methods based on passively monitoring scanning behavior on local IP addresses. We further explain in a mathematically rigorous fashion why detection techniques that are based only on this passive monitoring local IP addresses cannot quickly detect worm epidemics with a low false alarm rate.

The worm epidemic detection methods discussed in this paper are developed from the field of detection and estimation theory which has been classically applied to the signal processing of radar systems and to fault detection in manufacturing systems. A brief introduction to this field is given, but a more in-depth review of this material is provided by Poor [1994]. Of particular concern to the material in this paper is the notion of sequential change detection which is discussed more deeply by Basseville and Nikiforov [1993] and Wald [1947].

Sequential hypothesis testing methods for detecting malicious port-scanning has been proposed by Jung et al. [2004], Schechter et al. [2004] and Weaver et al. [2004]. The Threshold Random Walk (TRW) sequential hypothesis testing methods for detecting malicious port-scanning has been proposed by Jung et al. [2004], but the problem of RCS epidemic detection or the limitations of sequential hypothesis testing as discussed in this paper has not been considered by Jung et al. [2004]. Versions of the TRW method have been used by Schechter et al. [2004] and Weaver et al. [2004] for the detection and containment of scanning worms, respectively. Both Schechter et al. [2004] and Weaver et al. [2004] have focused on the problem of detecting infected hosts in a local network rather than the existence of an RCS worm epidemic in the global Internet, which is one of the major foci of this paper. Additionally, neither Jung et al. [2004] nor Weaver et al. [2004] have discussed the limitations of their scanning worm detection methods.

There have been several approaches to the automated anomaly-based detection of RCS Internet worms, most notably by Zou et al. [2003] where a Kalman filtering method is used in conjunction with observations based on ingress filters on routers for local networks. An important contribution of [Zou et al. 2003] is that the Kalman filtering method is used to determine an RCS worm epidemic's infection rate when a worm epidemic is propagating. However, no discussion is given in [Zou et al. 2003] on how the false rates are related to the detection speeds through the adjustment of Kalman filtering method's parameters. Similar to [Zou et al. 2003], the detection method discussed in this paper is designed to be

implemented on a local level, so that the automated epidemic detection policies operate by making observations of scanning behavior on unused local network addresses.

There is an inherent level of "background radiation" of undesired scanning that occurs even in unused addresses in local networks connected to the Internet [Pang et al. 2004] which can hinder the ability of anomaly-based detection methods to diagnose the existence of a worm epidemic in its early stages of growth. Fortunately the observed noise levels on unused network addresses are generally at least as low as the noise levels on active addresses, which aids the detection methods discussed below. In order to simplify the worm propagation models used in this paper, we assume that when a host attempts to scan and propagate a worm epidemic there is a uniform distribution of any one Internet address of being scanned, but this restriction can be easily removed.

The paper is organized as follows. Section 2 establishes the notation used throughout the paper and presents the well-known deterministic simple epidemic model. A density-dependent Markov jump model for worm propagation is introduced in Section 3. Section 4 presents a hybrid deterministic/stochastic point process model for a worm's scanning behavior as observed on a local network. Section 5 discusses models for the traffic observed on a local network during an RCS worm epidemic that incorporates both hybrid scanning model of the worm and background noise. Section 6 provides a brief introduction to the field of hypothesis testing and presents an optimal sequential hypothesis-testing method for the detection of scanning worm epidemics under idealized conditions. Section 7 introduces an anomaly-based Sequential Probability Ratio Test (SPRT) optimal detection method for RCS worm epidemics. Section 8 discusses fundamental limitations for this anomaly-based RCS worm epidemic testing method under idealized conditions. This shows that there are inherent limitations to the usefulness of anomaly-based epidemics detection methods for RCS worms. The paper concludes with a discussion of the results and possible areas for future research in Section 9. This paper is an extended and combined journal version of two initial conference papers [Rohloff and Başar 2005b; 2005a].

2. RANDOM CONSTANT SCANNING WORM PROPAGATION

It is assumed that an RCS worm can propagate over a network (such as the Internet) with n unique hosts. Of these addresses, $n_s \leq n$ hosts could potentially become infected by the worm. At a given time $t \geq 0$, the set of n_s potential hosts is split into *infected* and *susceptible* subpopulations, represented by $I(t)$ and $S(t)$ respectively. $I(t)$ is the number of hosts which are infected by the worm at time t, and $S(t)$ is the number of hosts which could become infected, but are not at time t. At time 0, $(S(0), I(0)) = (s_0, i_0)$ where $i_0 \geq 1$ is the initial infected population. Due to the random scanning propagation behavior of RCS worms, at $t > 0$, $S(t)$ and $I(t)$ are random variables.

Infected hosts propagate their infection by sending infectious packets to other randomly selected hosts in the general population at a constant rate. If a host that is uninfected but susceptible to infection receives an infectious packet, then that host becomes infected. The process of an infected host sending out infectious packets is called scanning. For general simplicity in the model it is assumed that infectious hosts are not removed from the general population so that for all $t \geq 0$, $I(t) + S(t) = n_s$. However, the results of this paper readily generalize to the cases of host recovery and/or removal.

When an infected host attempts to spread the worm by scanning randomly selected hosts, the addresses selected for scanning are assumed to be selected with a uniform distribution.

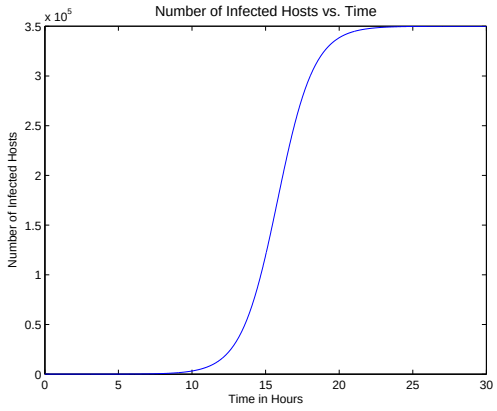

Fig. 1. Plot of a Deterministic CodeRed1v2 Propagation Simulation.

(This assumption is made for mathematical simplicity, but can be removed.) Therefore, any one infectious packet sent at time t has probability $\frac{S(t)}{n}$ of being sent to a susceptible host. It is assumed that an infected host scans for susceptible hosts at a constant rate β which is called the *infection parameter*. Therefore $\frac{\beta}{n}S(t)$ is the rate at which an infected host transmits its infection to susceptible hosts and $\frac{\beta}{n}S(t)I(t)$ is the rate at which the infected population transmits the infection to susceptible hosts.

Now, for the deterministic simple epidemic model, the variable $s(t)$ is used to represent the size of the susceptible population at time t, and the variable $i(t)$ is a similarly defined deterministic variable which represents the number of infected individuals at time t. Consequently, $\frac{\beta}{n}s(t)i(t)$ is the rate at which the $i(t)$ infected hosts propagate the epidemic, and

$$\frac{di}{dt} = \frac{\beta}{n}s(t)i(t) = \frac{\beta}{n}(n_s - i(t))i(t).$$

With $i(0) = i_0$,

$$i(t) = \frac{i_0 n_s}{i_0 + (n_s - i_0)e^{-\beta\frac{n_s}{n}t}}.$$

This deterministic Kermack-McKendrick model of an epidemic propagation, originally proposed by Kermack and McKendrick [1927], is an approximation of the underlying stochastic process $(S(t), I(t))$. The use of this deterministic approximation model in the context of RCS worms has yet to be justified in a mathematically rigorous fashion. In Section 3 below, we present a stochastic propagation model of an RCS worm epidemic derived from first principles and discuss conditions under which the simple epidemic model is a sufficient approximation for the RCS epidemic process.

A plot of $i(t)$ versus t, called the deterministic *infection curve*, for a model of the CodeRed1v2 epidemic can be seen in Figure 1 where it is assumed that $n = 2^{32}$ (the IP address space), $n_s = 350,000$ (an approximation of the size of the susceptible CodeRed population), $\beta = 10188$ (an approximation of the number of IP addresses scanned by an infected host scans per hour) and $i_0 = 1$ (the size of the initial infection).

3. STOCHASTIC EPIDEMIOLOGICAL MODEL FOR SCANNING WORMS

The deterministic Kermack-McKendrick epidemic model presented above is a deterministic abstraction of a process that is inherently stochastic. We now present a stochastic density-dependent Markov jump process propagation model for an RCS worm drawn from the field of epidemiology [Andersson and Britton 2000; Daley and Gani 1999], but not previously discussed in the computer worm literature.

When $(S(t), I(t)) = (s, i)$, the pair (s, i) is the "state" of the epidemic. Due to the propagation of the RCS worm infection and the assumption that $S(t) + I(t) = n_s$, if the propagation process is at a state (s, i), then the next state must be $(s-1, i+1)$ and the next state after that must be $(s-2, i+2)$ and so on until state $(0, n_s)$ is reached. From $(0, n_s)$ no other state can be reached, so $(0, n_s)$ is an absorbing state and almost surely a time t^{fin} is eventually reached such that $(S(t^{fin}), I(t^{fin})) = (0, n_s)$ [Hoel et al. 1971].

Because the destinations of the infectious packets are selected by the infectious hosts with a uniform distribution, any one infectious packet sent at time t has a probability of $\frac{S(t)}{n}$ of being sent to a susceptible host. At time t, the $I(t)$ infectious hosts transmit infectious packets each at the rate β, so $\frac{\beta}{n} S(t) I(t)$ is the rate at which $(S(t), I(t)) = (s, i)$ goes to $(s-1, i+1)$. From Daley and Gani [1999], this process can then be modeled as a jump process with a jump intensity:

$$q_{(s_a, i_a)(s_b, i_b)} = \begin{cases} \frac{\beta}{n} s_a i_a & \text{if } (s_b = s_a - 1) \wedge (i_b = i_a + 1) \\ 0 & \text{otherwise} \end{cases} .$$

This jump process is Markovian because at state $(S(t), I(t)) = (s, i)$, the current jump intensity depends only on the current state (s, i) and is independent of the previous states of the process. Consequently, this stochastic epidemic propagation process is by definition a *density-dependent Markov jump process* because the jump intensity at state (s, i) depends on the "densities" of the number of susceptible hosts s and the number of infected hosts i. Several important aspects of this subclass of Markov jump processes are discussed by Andersson and Britton [2000] and Ethier and Kurtz [1986].

Five different simulations of a stochastic worm propagation model with growth parameters similar to that of the CodeRed1v2 worm and an initial infection of $i_0 = 1$ can be seen in Figure 2. Note that by visual inspection, the propagation curves in Figure 2 are approximately the same curve shifted in time.

It has been shown by Andersson and Britton [2000] that the expected values of the susceptible and infected population sizes in the stochastic model at time t, $(E\{S(t)\}, E\{I(t)\})$, converge almost surely to the susceptible and infected population sizes predicted by the deterministic model $(s(t), i(t))$ as the size of the populations n_s and n increase. Also, the fluctuations of the susceptible and infected population sizes in the stochastic model around the deterministic solution are asymptotically Gaussian. However, as discussed by Andersson and Britton [2000], it is difficult if not impossible to find a closed-form expression for the covariance of a density-dependent Markov jump process. Despite this, the covariances of worm propagation at various time intervals can be easily computed through simulation. The mean of 100 such CodeRed1v2 epidemic propagation simulations and the variance of these simulations at various instances of time can be seen in Figure 3 where initially one host is infected.

As predicted by the plots in Figure 3, $(E\{S(t)\}, E\{I(t)\})$ is approximately the same as the deterministic simulation $(s(t), i(t))$ seen in Figure 1, but the variance of the stochastic

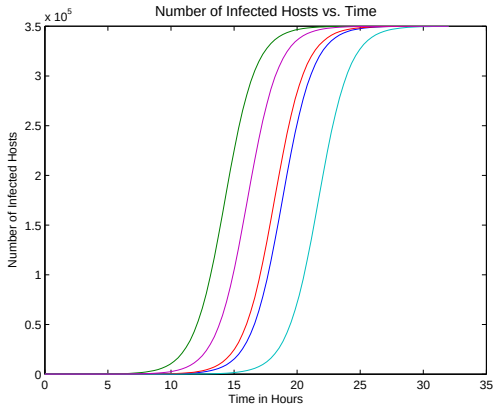

Fig. 2. Plot of Five Stochastic CodeRed1v2 Propagation Simulations.

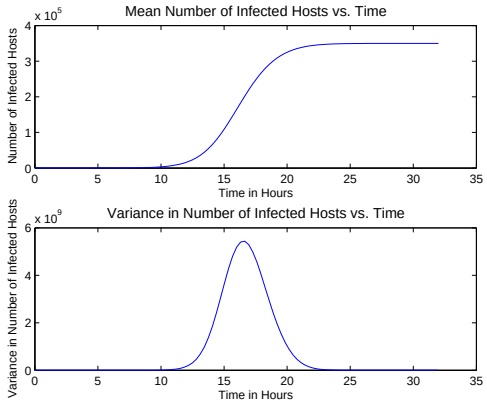

Fig. 3. Plots of the Mean and Variance of 100 Stochastic CodeRed1v2 Propagation Simulations.

epidemic simulations is potentially very large. Consider also the first plot of Figure 4 which shows five stochastic simulations of the propagation of a worm with growth parameters similar to that of the CodeRed1v2 worm where initially half of the susceptible population is infected. The variances computed from 100 such simulations can be seen in the second plot.

From the first plot of Figure 4, the various simulations of the epidemic propagations are effectively identical. This is also indicated in the variance plot of Figure 4 where the maximum variance of these simulations (where initially half of the susceptible population is infected) is several orders of magnitude less than the maximum variance of the simulations in Figure 3 (where initially one host is infected). This indicates that the epidemic propagation curves of the stochastic epidemic simulations in Figure 2 are effectively all identical but shifted in time. However, although the propagation curves seen in Figure 2 are nearly identical, there can be differences on the order of hours in the amount of time it takes for a

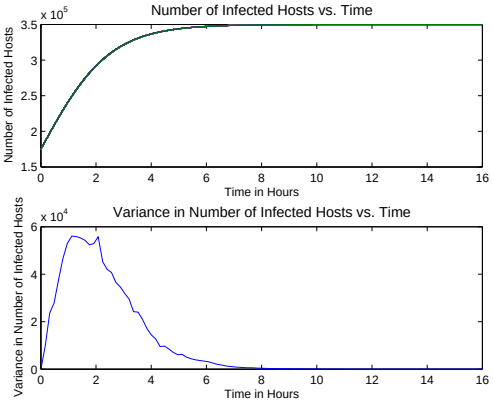

Fig. 4. Plots of 5 Stochastic CodeRed1v2 Propagation Simulations and the Variance of 100 Stochastic CodeRed1v2 Propagation Simulations.

worm epidemic to infect half of the susceptible population.

Define T_{i_j} to be a random variable which represents the amount of time the infection process is in state (s_j, i_j). Every scanning attempt by an infected host during the epidemic is a Bernoulli trial with probability s_j/n of successfully scanning a host in the set of s_j susceptible hosts. Because these Bernoulli trials occur at the rate of βi_j, T_{i_j} is exponentially distributed with mean $\frac{n}{\beta(n_s-i_j)i_j}$ and variance $(\frac{n}{\beta(n_s-i_j)i_j})^2$ [Stark and Woods 1994]. Furthermore, $T_1, \ldots, T_{n_s}$ are all independent because the stochastic epidemic propagation process is Markovian. Note that $\max_{i_j}(Var(T_{i_j})) = \left(\frac{n}{\beta}\right)^2 \left(\frac{1}{(n_s-1)^2}\right)$ and $\min_{i_j}(Var(T_{i_j})) = \left(\frac{n}{\beta}\right)^2 \left(\frac{4}{n_s^4}\right)$.

Let $i_a, i_b \in \{1, \ldots, n_s\}$ be such that $i_b > i_a$ and let $T_{i_a i_b}$ be a random variable that represents the amount of time it takes a stochastic infection propagation process to go from state (s_a, i_a) to state (s_b, i_b). By definition, $T_{i_a i_b} = \sum_{i_j=i_a}^{i_b-1} T_{i_j}$. In general, $T_{i_a i_b}$ is not exponentially distributed, but a closed-form expression for its probability distribution function exists. Relevant to the discussions in this paper,

$$\begin{aligned} E\{T_{i_a i_b}\} &= E\{\sum_{i_j=i_a}^{i_b-1} T_{i_j}\} \\ &= \sum_{i_j=i_a}^{i_b-1} E\{T_{i_j}\} \\ &= \sum_{i_j=i_a}^{i_b-1} \frac{n}{\beta} \frac{1}{(n_s-i_j)i_j} \\ &= \frac{n}{\beta} \sum_{i_j=i_a}^{i_b-1} \frac{1}{(n_s-i_j)i_j} \end{aligned} \quad (1)$$

and

$$\begin{aligned} Var(T_{i_a i_b}) &= Var\left(\sum_{i_j=i_a}^{i_b-1} T_{i_j}\right) \\ &= \sum_{i_j=i_a}^{i_b-1} Var(T_{i_j}) \\ &= \sum_{i_j=i_a}^{i_b-1} \left(\frac{n}{\beta}\right)^2 \left(\frac{1}{(n_s-i_j)i_j}\right)^2. \\ &= \left(\frac{n}{\beta}\right)^2 \sum_{i_j=i_a}^{i_b-1} \left(\frac{1}{(n_s-i_j)i_j}\right)^2. \end{aligned}$$

It can also be shown that $E\{T_{1\frac{n_s}{2}}\} = \frac{n}{\beta n_s}[C+\ln(n_s-1)+f(n_s)]$ where $C \neq 0$ is a constant and $f(n_s) \in O(1/n_s)$.

From Equation 1,

$$E\{T_{1\frac{n_s}{2}}\} = \frac{n}{\beta}\sum_{i_j=1}^{\frac{n_s}{2}-1} \frac{1}{(n_s-i_j)i_j}.$$

Therefore,

$$\begin{aligned} E\{T_{1\frac{n_s}{2}}\} &= \frac{n}{\beta}\sum_{i_j=1}^{\frac{n_s}{2}-1} \frac{1}{(n_s-i_j)i_j} \\ &= \frac{n}{\beta}\left(\sum_{i_j=1}^{\frac{n_s}{2}-1} \frac{\frac{1}{n_s}}{n_s-i_j} + \sum_{i_j=1}^{\frac{n_s}{2}-1} \frac{\frac{1}{n_s}}{i_j}\right) \\ &= \frac{n}{\beta n_s}\left(\sum_{i_j=1}^{n_s-1} \frac{1}{i_j} - \sum_{i_j=1}^{\frac{n_s}{2}} \frac{1}{i_j} + \sum_{i_j=1}^{\frac{n_s}{2}-1} \frac{1}{i_j}\right) \\ &= \frac{n}{\beta n_s}\left(\sum_{i_j=1}^{n_s-1} \frac{1}{i_j} - \frac{2}{n_s}\right) \end{aligned}$$

Then, by Equation 0.131 from Gradshteyn and Ryzhik [1994],

$$E\{T_{1\frac{n_s}{2}}\} = \frac{n}{\beta n_s}[C+\ln(n_s-1)+f(n_s)]. \tag{2}$$

Note that this expression for $E\{T_{1\frac{n_s}{2}}\}$ is not equal to the expression $\frac{n}{\beta n_s}\ln(n_s-1)$. That is, the expected time for a worm epidemic to infect half of the susceptible hosts predicted by the stochastic model with an initial infected population of 1 is not equal to the same value predicted by the deterministic model in Section 2. More quantitatively, for the epidemic parameters used to generate the simulations of the CodeRed1v2 worm in the plots shown above, $E\{T_{1\frac{n_s}{2}}\}$ differs from $\frac{n}{\beta n_s}\ln\frac{n_s-1}{1}$ by over half an hour. This is due to the fact that both the deterministic and stochastic models are different abstractions of the same underlying process where the deterministic model assumes a continuous state process, which is not a

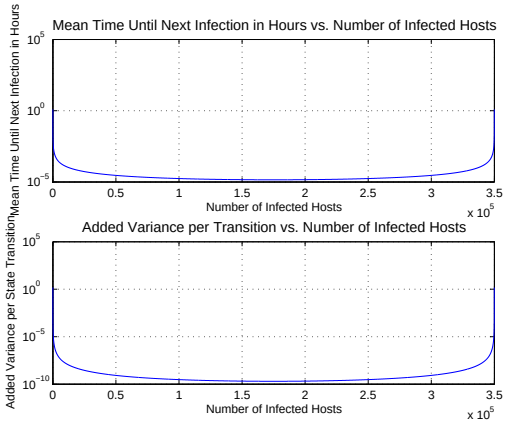

Fig. 5. Plot of the mean and variance of the inter-state jump intervals for a CodeRed1v2 worm.

completely accurate reflection of the underlying discrete behavior of a worm's propagation. However, as n_s increases, $C\frac{n}{\beta n_s} \ll \ln(n_s-1)\frac{n}{\beta n_s}$ and $f(n_s)\frac{n}{\beta n_s} \ll \ln(n_s-1)\frac{n}{\beta n_s}$. Therefore, as n_s increases, $E\{T_{1\frac{n_s}{2}}\} = \frac{n}{\beta n_s}[C+\ln(n_s-1)+f(n_s)]$ approaches $\frac{n}{\beta n_s}\ln(n_s-1)$.

Figure 5 contains plots of $E\{T_{i_j}\}$ versus i_j and $Var(T_{i_j})$ versus i_j on semilog scales. Both of the plots in Figure 5 are bowl-shaped with relatively flat bottoms with steep sides. Just after initialization and $I(t)$ is close to 0, there is potentially large variation in the evolution of $I(t)$ as indicated in Figure 5 (hence the observed time-shifting in the multiple simulated growth curves in Figure 5.) Conversely, once the epidemic has been established on a large enough portion of the susceptible population and there is still a relatively large number of susceptible hosts left to infect, the epidemic's growth pattern is relatively stable and the simulated epidemic growth curves in Figure 2 are nearly identical (except for time-shifting).

For an RCS worm epidemic to be a threat to the Internet, n_s, the size of the susceptible population, should be relatively large (on the order of several thousand or more hosts). If this holds then n_s^4 will be very large compared to $(n_s-1)^2$ and hence $\min_{i_j}(Var(T_{i_j}))$ will be much smaller than $\max_{i_j}(Var(T_{i_j}))$. This indicates that $Var_{T_{i_a i_b}}$ will be relatively small when $0 << i_a$ and $i_b << n_s$ compared to $Var_{T_{i_a i_b}}$ when i_a is close to 0 or i_b is close to n_s. This explains why the propagation simulation curves in Figure 2 appear to be the same curve shifted in time.

This analysis of $Var_{T_{i_a i_b}}$ indicates that excluding the earliest and latest growth stages of an epidemic and the "time-shifting" effects due to the probability distribution on $T_{i_a i_b}$, the simulations of a worm's propagation using the deterministic and stochastic models are effectively equivalent when n_s is sufficiently large. Furthermore, if n_t is fairly small, the probability of actually seeing *any* effects due to worm propagation during the earliest stages of an RCS worm epidemic is very small (when propagation variability is at its greatest), even if $Var_{T_{i_a i_b}}$ is very large during these early stages. A relatively small value for n_t implies that a network administrator has implemented a network architecture where the ratio of used network addresses to fallow addresses instrumented and available for worm detection is very large, which is most often the case. The possible variability in the

stochastic propagation of RCS worm epidemics mentioned by Moore et al. [2003], Zou et al. [2003] and Nicol [2006] is surprisingly minor under these conditions, and our assumption of these conditions justifies our use of the simple epidemic model from Kermack and McKendrick [1927] as a reasonable approximation of the density-dependent Markov process model derived from first principles in the remainder of the paper.

4. OBSERVED EPIDEMIC AND RADIATION LOCAL SCANNING MODEL

Based on the analyses of the deterministic and stochastic models for a worm's global propagation above, we introduce a new hybrid model for the scanning behavior of a global RCS worm epidemic observed on a local network. Let random variables $\tau_1^w, \tau_2^w, \tau_3^w, \ldots$ represent the times at which scans due to the worm are observed on the local network. For computational simplicity this hybrid model uses the deterministic simple epidemic model to describe the large-scale global propagation behavior, but a stochastic model to generate $\tau_1^w, \tau_2^w, \tau_3^w, \ldots$ based on the current state of the large-scale propagation.

Let n_t be the number of unique addresses in the local network over which scanning observations are made. Because it is assumed that the addresses selected for scanning by infected hosts are assumed to be selected with a uniform distribution, every scanning attempt by an infected host during the epidemic is a Bernoulli trial with probability n_t/n of successfully scanning a host in the set of n_t local addresses. Let $\Phi_1, \Phi_2, \Phi_3, \ldots$ be random variables which for $i \in \{1, 2, \ldots\}$, Φ_i represents the global number of scans which have occurred due to the worm up until the time of the ith successful scan of the local network. Because each scan is a Bernoulli trial with probability n_t/n of success, $\Phi_1, \Phi_2 - \Phi_1, \Phi_3 - \Phi_2, \ldots$ are independent and geometrically distributed random variables, all with mean n/n_t. Because n_t is small compared to n, $\Phi_1, \Phi_2 - \Phi_1, \Phi_3 - \Phi_2, \ldots$ can be approximated as being exponentially distributed with mean n/n_t.

We now show how $\Phi_1, \Phi_2 - \Phi_1, \Phi_3 - \Phi_2, \ldots$ can be used to find $\tau_1^w, \tau_2^w, \tau_3^w, \ldots$: the times at which scans due to the worm are observed on the local network. Let $\Phi(t)$ be the random variable which represents the total number of global infection attempts that have been made due to the epidemic propagation behavior up to time t using the stochastic model, and let $\phi(t)$ be its deterministic approximation. Then,

$$\frac{d\phi}{dt} = \beta i.$$

It is already known from the dynamics of the deterministic epidemic model that

$$\frac{di}{dt} = \frac{\beta}{n} si \tag{3}$$

where $s = n_s - i$. Therefore, by basic mathematical manipulation,

$$\frac{n}{(n_s - i)} di = d\phi.$$

Consequently,

$$\int_{i_0}^{i(\phi)} \frac{n}{(n_s - i)} di = \int_0^{\phi} d\phi.$$

With basic mathematical manipulation,

$$i(\phi) = n_s - \frac{n_s - i_0}{e^{\phi/n}}. \tag{4}$$

By substituting Equation 4 into Equation 3,

$$\begin{aligned}\frac{di}{dt} &= \frac{\beta}{n}\left(\frac{n_s - i_0}{e^{\phi/n}}\right)\left(n_s - \frac{n_s - i_0}{e^{\phi/n}}\right) \\ &= \frac{\beta}{n}\left(\frac{n_s - i_0}{e^{\phi/n}}\right)\left(n_s - (n_s - i_0)e^{-\phi/n}\right)\end{aligned} \tag{5}$$

and from equation 4,

$$\frac{di}{d\phi} = \frac{n_s - i_0}{n}e^{\phi/n}. \tag{6}$$

Therefore,

$$\frac{dt}{d\phi} = \frac{1}{\beta\left[n_s - (n_s - i_0)e^{-\phi/n}\right]}.$$

Finally,

$$t(\phi) = \frac{n}{\beta n_s}\left[\frac{\phi}{n} + \ln\left(\frac{n_s - (n_s - i_0)e^{-\phi/n}}{i_0}\right)\right]. \tag{7}$$

If ϕ scans have occurred due to the worm globally, then $t(\phi)$ represents the time at which the ϕth scan occurred under the assumption of epidemic propagation dynamics modeled by the simple epidemic model.

With Equation 7, one can also obtain $\phi(t)$ which is the number of scans which have occurred due solely to the propagation of the worm epidemic at time t:

$$\phi(t) = \beta n_s t + n\ln\left[\frac{i_0 + (n_s - i_0)e^{-\beta\frac{n_s}{n}t}}{n_s}\right]. \tag{8}$$

With the above equations, one can approximate $\{\tau_1^w, \tau_2^w, \tau_3^w, \ldots\}$ as $\{t(\Phi_1), t(\Phi_2), t(\Phi_3), \ldots\}$ when one can accurately model the RCS worm epidemic propagation process using the deterministic model.

With a closed-form expression for $t(\phi)$, a model can now be defined to generate the times $\tau_1^w, \tau_2^w, \tau_3^w, \ldots$ at which scans due to the worm are observed on the local network. Recall that $\Phi_1, \Phi_2 - \Phi_1, \Phi_3 - \Phi_2, \ldots$ are independently exponentially distributed and can therefore be generated using standard methods during simulation. Hence, values for $\Phi_1, \Phi_2 - \Phi_1, \Phi_3 - \Phi_2, \ldots$ can be used to generate $\Phi_1, \Phi_2, \Phi_3, \ldots$ during simulation. Then, using the $t(\phi)$ function, $\tau_1^w = t(\Phi_1), \tau_2^w = t(\Phi_2), \tau_3^w = t(\Phi_3), \ldots$. An example of $\tau_1^w, \tau_2^w, \tau_3^w, \ldots$ values generated using a Matlab simulation of the CodeRed1v2 model with $n_t = 15$ can be seen in Figure 6.

We describe next how this hybrid model can be justified as an accurate representation of the locally observed behavior due to a worm epidemic when the variability in the initial propagation of a worm described in Section 3 can be ignored.

If it is assumed that $n_s >> i_0$, then

$$i(\phi) = n_s\left(1 - \frac{1}{e^{\phi/n}}\right).$$

Also,

$$E\{\Phi_k\} = E\{\Phi_k - \Phi_{k-1} + \Phi_{k-1} - \Phi_{k-2} + \Phi_{k-2} - \cdots - \Phi_1 + \Phi_1\}.$$

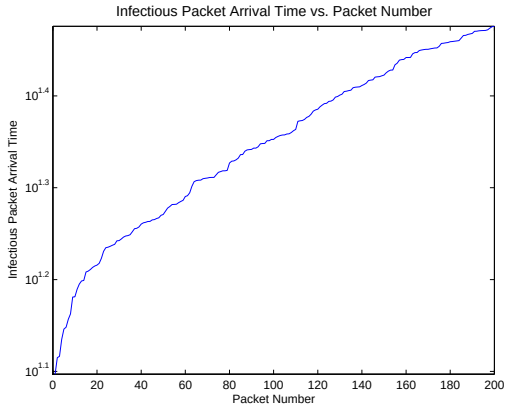

Fig. 6. Plot of Infectious Packet Arrival Times for a CodeRed1v2 worm with $n_t = 15$.

It is known that $\{(\Phi_k - \Phi_{k-1}), (\Phi_{k-1} - \Phi_{k-2}), \cdots, (\Phi_2 - \Phi_1), \Phi_1\}$ are all independent, identically distributed with mean n_t/n. Therefore,

$$\begin{aligned} E\{\Phi_k\} &= E\{\Phi_k - \Phi_{k-1}\} + E\{\Phi_{k-1} - \Phi_{k-2}\} \cdots E\{\Phi_2 - \Phi_1\} + E\{\Phi_1\}\} \\ &= kn_t/n. \end{aligned}$$

Consequently,

$$\begin{aligned} i(E\{\Phi_k\}) &= n_s \left(1 - \frac{1}{e^{\phi(E\{\Phi_k\})/n}}\right) \\ &= n_s \left(1 - \frac{1}{e^{\frac{1}{kn_t}}}\right). \end{aligned}$$

Recall that $T_{i(E\{\Phi_k\})}$ is the amount of time it would take a stochastic epidemic process to transition to the next infection state from $i(E\{\Phi_k\})$, the size of the global infection when the kth infectious packet is expected to be observed locally. Substituting $i(E\{\Phi_k\})$ into the equation for $Var(T_{i_j})$,

$$Var(T_{i(E\{\Phi_k\})}) = \left(\frac{n}{\beta}\right)^2 \frac{1}{n_s^4 \left(e^{-\frac{1}{kn_t}}(1 - e^{-\frac{1}{kn_t}})\right)^2}$$

Recall that $Var(T_{i(E\{\Phi_k\})})$ represents the variance in the amount of time it takes the stochastic epidemic propagation process to jump from state $i(E\{\Phi_k\})$ to $i(E\{\Phi_k\}) + 1$. Therefore, $Var(T_{i(E\{\Phi_k\})})$ gives an indication of the stability of the epidemic's propagation when the kth scan due to the worm is observed on the local network. If $Var(T_{i(E\{\Phi_1\})})$ is relatively small, then the epidemic has been suffi ciently established on the global network such that the deterministic simple epidemic model will accurately represent its future large-scale growth after the fi rst scanning packet is observed on the local network. Recall that $\max_{i_j}(Var(T_{i_j})) = \left(\frac{n}{\beta}\right)^2 \left(\frac{1}{(n_s-1)^2}\right)$ and $\min_{i_j}(Var(T_{i_j})) = \left(\frac{n}{\beta}\right)^2 \left(\frac{4}{n_s^4}\right)$. Therefore, if n_s is

relatively large or n_t is relatively small, as is generally assumed for realistic worms and network conditions, then $Var(T_{i(E\{\Phi_1\})})$ is on the same order of magnitude as $\min_{i_j}(Var(T_{i_j}))$. Hence, if n_s is relatively large or n_t is relatively small, then the first infectious packet is expected to be sent to an address on the local network when the underlying stochastic worm epidemic process is in a relatively stable region of growth with little variance in the growth dynamics. This also means that there is unlikely to be any observations of scanning behavior on the local network when the deterministic model is not as good a representation of the worm's growth dynamics. Therefore, if the exact starting time of the worm propagation is not known, then the deterministic epidemic model is an effective representation of the stochastic epidemic process under these conditions, and the hybrid observation model would likewise be an effective representation of the local observation process $\tau_1^w, \tau_2^w, \tau_3^w, \ldots$ when n_t is small and n_s is large. This therefore motivates both when and why the deterministic model can be used as a simple, but accurate approximation of the underlying RCS worm stochastic epidemic propagation process.

5. OBSERVED EPIDEMIC LOCAL SCANNING MODEL

As was discussed by Pang et al. [2004], not all of the scanning behavior observed on a set of locally unassigned addresses in a local network is solely due to the active propagation of a worm epidemic. Pang et al. [2004] show that the traffic on a local network is inherently noisy with a substantial volume of "background radiation": undesired scanning behavior on a local network not due to the propagation of a worm. Therefore, if one is attempting to detect the existence of an RCS worm epidemic by solely observing the scanning traffic on a set of unused local addresses, the observation of scanning traffic on an unused IP address is not an immediate indicator of the existence of an Internet worm. With the stochastic model for the scanning behavior of an RCS worm epidemic process as observed on a relatively small set of network addresses, we develop models for the scanning behavior observed on an unused address space in a local network both under normal conditions (when the observed scanning traffic is due solely to background radiation) and when an RCS worm epidemic is propagating (when the observed scanning traffic is due to radiation and the epidemic propagation).

Although we have discussed how to model observations during a worm epidemic due solely to the epidemic's propagation behavior, our previous model did not account for the randomized, non-trivial background radiation. The interarrival times between background radiation packets observations on a local network are modelled here as being independently and exponentially distributed with parameter η. The value of η has been found to be heavily dependent on the day of the week and several other factors such as the type of packets being observed (TCP versus UDP). However, for mathematical simplicity we assume that the value of η is constant with respect to the amount of time it takes an RCS worm to propagate. By visual inspection of the data shown from Pang et al. [2004], approximately 5 TCP background radiation packets hit a given IP address per hour on the Internet on average, and about 0.2 radiation packets per IP address per hour are observed for UDP packets. A simulation of the UDP packet background radiation scan arrival times $\{\tau_1^n, \tau_2^n, \tau_3^n, \ldots\}$ observed on a local address space with $n_t = 15$ can be seen in Figure 7 when $\eta = 0.2$.

Now, with a mathematical model for $\{\tau_1^w, \tau_2^w, \tau_3^w, \ldots\}$ for a given worm and $\{\tau_1^n, \tau_2^n, \tau_3^n, \ldots\}$ for a given level of background radiation, a set

$$\{\tau_1, \tau_2, \tau_3, \ldots\} = \{\tau_1^w, \tau_2^w, \tau_3^w, \ldots\} \cup \{\tau_1^n, \tau_2^n, \tau_3^n, \ldots\}$$

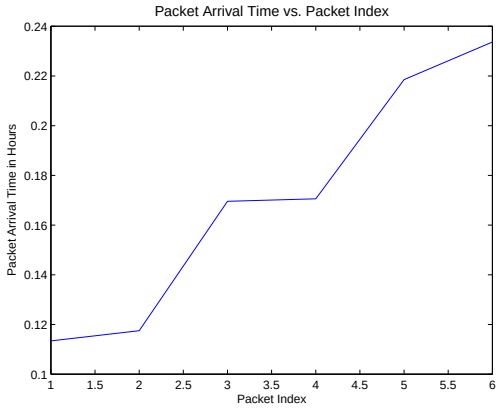

Fig. 7. Plot of Scanning Times on Local Network Due to UDP Background Radiation when $\eta = 0.2$.

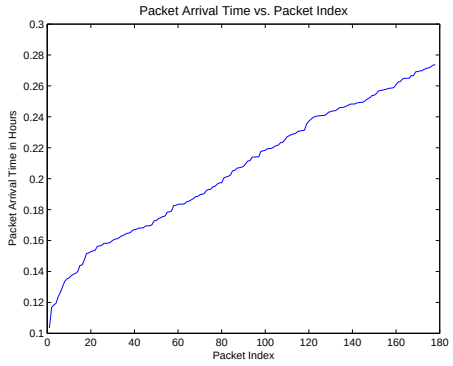

Fig. 8. Plot of Scanning Times on Local Network Due to a Simulated Slammer Outbreak and Background Simulation.

can be defined where $\{\tau_1, \tau_2, \tau_3, \ldots\}$ are the ordered times of all scans of the local address space such that τ_i is the time of the arrival of the ith packet in unused address space of the local network. A simulation of this set of composed scan arrival times $\{\tau_1, \tau_2, \tau_3, \ldots\}$ observed on a local address space with $n_t = 15$ can be seen in Figure 8.

6. HYPOTHESIS TESTING FOR WORM EPIDEMICS

In this section we present an optimal Sequential Probability Ratio Test (SPRT) for the detection of a worm epidemic propagating over the Internet under idealized conditions. A brief introduction to SPRT's is given here, but a more extensive overview is provided in Chapter 4 from Basseville and Nikiforov [1993].

In the framework of simple hypothesis testing it is assumed that there are two possible hypotheses to describe a system, H_0 and H_1. These hypotheses correspond to two (possibly non-stationary) probability distributions (Φ_0 and Φ_1 respectively) on observations $Y^n =$

$\{o_1, o_2, o_3, \ldots\}$ that could be made of the system. Based on the set $Y^n = \{o_1, o_2, o_3, \ldots\}$ of observations of system behavior, it should be decided whether hypothesis H_0 or H_1 better describes the system. In the context of the worm epidemic detection problem, given a set of observations of local scanning times $\{\tau_1, \tau_2, \tau_3, \ldots\}$, the hypothesis testing problem is to decide whether those observed packet arrivals were caused in part by a worm epidemic. Hypothesis H_0 is said to hold if solely background radiation scanning caused the observations $\{\tau_1, \tau_2, \tau_3, \ldots\}$, while hypothesis H_1 is said to hold if the scanning behavior due to a worm combined with background radiation scanning caused the observations $\{\tau_1, \tau_2, \tau_3, \ldots\}$.

Suppose a set of n observations $Y^n = \{o_1, o_2, \ldots, o_n\}$ of a system with two possible hypotheses H_0 and H_1 are given. Let $g(Y^n)$ be a hypothesis decision function for the observations Y^n such that if $g(Y^n) = 1$, then hypothesis H_1 is chosen to be the hypothesis of the current state of the system. Conversely, if $g(Y^n) = 0$, then hypothesis H_0 is chosen to be the hypothesis of the current state of the system. (In general it is possible that $g(Y^n)$ may be undefined for some input in order to indicate that neither hypothesis is chosen and more data may be necessary to select a hypothesis.) For an observation Y^n, define $p_0(Y^n)$ to be the probability density of Y^n when hypothesis H_0 holds, and define $p_1(Y^n)$ to be the probability density of Y^n when hypothesis H_1 holds. If

$$S_n = \ln\left(\frac{p_1(Y^n)}{p_0(Y^n)}\right),$$

a decision function $g(Y^n)$ can be defined such that

$$g(Y^n) = \begin{cases} 1 & \text{if } S_n \geq h \\ 0 & \text{if } S_n \leq -a \\ \text{undefined} & \text{otherwise} \end{cases}. \tag{9}$$

Because S_n is a logarithm of a probability ratio, this decision function $g(\cdot)$ is called a probability ratio test where h and a are boundaries (thresholds) on the hypothesis decisions such that $-a \leq h$. It may be possible based on an observation Y^n to have a false alarm such that $g(Y^n) = 0$ when H_1 holds or $g(Y^n) = 1$ when H_0 holds. Note that if $-a < S_n < h$, then $g(Y^n)$ is undefined and neither hypothesis is chosen.

Sometimes it may not always be possible or desirable to have a fixed number of observations Y_n in order to decide between H_0 and H_1. For instance, for the RCS worm detection problem, once a worm commences scanning addresses in a local network, the worm epidemic should be detected as soon as possible without regard to the number of observations made. To this end, *sequential analysis* has been developed as a theory for solving hypothesis testing problems when the sample size of the observations is not fixed a priori [Basseville and Nikiforov 1993]. That is, observations of the behavior of a system might be made in an online manner and it should be decided which hypothesis of the system state holds as soon as possible while maintaining a desirable performance level such as a sufficiently small false alarm rate.

For the above probability ratio test $g(\cdot)$, let there be a set of observations $Y = \{o_1, o_2, \ldots\}$. Define the random variable T such that

$$T = min\{n > 1 : g(Y^n) \in \{0, 1\}\}.$$

The random variable T represents the minimum number of observations of system behavior necessary in order to choose a hypothesis using the decision function g. For $Y =$

$\{o_1, o_2, \ldots\}$, define the *Sequential Probability Ratio Test (SPRT)* as follows:

$$g'(Y) = \begin{cases} 1 & \text{if } g(Y^T) = 1 \\ 0 & \text{if } g(Y^T) = 0 \end{cases} . \tag{10}$$

The sequential decision function $g'(\cdot)$ is sequential in that it selects a hypothesis using the minimum number of observations Y^T to determine a hypothesis with respect to the decision function $g(\cdot)$.

The random variable T is called the *stopping time* because it is the time such that when T observations have been made, it is no longer necessary to collect more data to assign a hypothesis after T. The Average Run Length (ARL) is the mean number of observations $E_\theta(T)$ necessary for testing the hypotheses to obtain a given error rate if hypothesis H_θ holds. The ARL is an important performance measure of sequential analysis methods. Also related to the ARL is the notion of two different classes of false alarms. The variables α_0 and α_1 are defined to be the error rates for a sequential analysis methods where α_0 is the rate at which H_1 is thought to be true when H_0 holds, and α_1 is the rate at which H_1 is thought to be true when H_0 holds. Generally, as the threshold levels $-a$ and h are adjusted for the probability ratio test, the ARL decreases as the false alarm rates increase.

The SPRT is known to be an optimal sequential decision method [Basseville and Nikiforov 1993]. Suppose due to the thresholds $-a, h$ used in $g(\cdot)$ to define $g'(\cdot)$ the ARL is $E_\theta(T)$ and false alarm rates are α_0 and α_1. Let there be another sequential decision function $\tilde{g}'(\cdot)$ with ARL $E_\theta(\tilde{T})$ and false alarm rates $\tilde{\alpha}_0$ and $\tilde{\alpha}_1$ such that $\tilde{\alpha}_0 \leq \alpha_0$ and $\tilde{\alpha}_1 \leq \alpha_1$. Then, $E_0(\tilde{T}) \geq E_0(T)$ and $E_1(\tilde{T}) \geq E_1(T)$.

7. AN SPRT FOR RCS WORM EPIDEMICS

We show, in this section, how the SPRT in Equation 10 can be used for the detection of RCS worm epidemics on the Internet. Our hypothesis testing method uses information about the parameters of the background radiation along with the exact parameters of the worm epidemic if H_1 were to hold (namely β, i_0 and n_s.) The worm epidemic propagation parameters will not be available in practice, but this information is used in the following section to demonstrate some fundamental performance limitations of optimal sequential analysis methods for the detection RCS worms.

Consider the observed scanning interarrival time data seen in Figure 9. The top set of data is a graph of simulated TCP interarrival times due to both background radiation and the propagation of an RCS worm with parameters the same as the simulated CodeRed1v2 worm in Section 2. The TCP background radiation simulation uses the parameters from [Pang et al. 2004]. The CodeRed1v2 propagation begins propagation at $t = 0$. For the first several hours, there is little or no observed scanning due to the worm. However, during the last several hours, the worm is fully propagated and scanning the local host at its peak rate. As can be seen from the simulation, it may not be immediately obvious that a worm is propagating by simply "looking" at scanning data.

Suppose a set of scanning packet arrival times $\{\tau_1, \tau_2, \ldots, \tau_n\}$ are observed on the local network. Let $\phi(t)$ be the function that maps a packet scanning time to its scanning index. If there is no worm on the Internet, let $\phi^0(t) = \eta t$ be the function which maps a packet arrival time to its scanning index, but if there is a worm on the Internet, with Equation 8 above,

$$\phi^1(t) \;=\; \left[\beta n_s t + n \ln\left[\frac{i_0 + (n_s - i_0)e^{\frac{-\beta n_s t}{n}}}{n_s}\right]\right] + \eta t$$

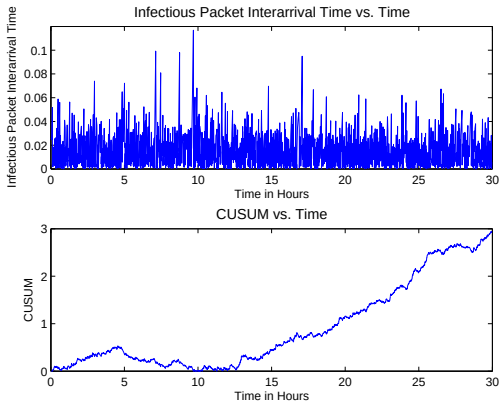

Fig. 9. Plots of Observed Scanning Data and Computed CUSUM Data for a CodeRed Epidemic Using SPRT.

$$= (\beta n_s t + \eta)t + n \ln \left[\frac{i_0 + (n_s - i_0)e^{\frac{-\beta n_s t}{n}}}{n_s} \right]$$

is the function which maps a packet arrival time to its scanning index. Hence, for a given set of local scanning time observations $\{\tau_1, \tau_2, \ldots, \tau_n\}$, two sets of packet indices can be computed for the two worm existence hypotheses H_0 and H_1. Define $\{\phi_1^0, \phi_2^0, \ldots, \phi_n^0\}$ to be the packet indices under the assumption of hypothesis H_0 using $\phi^0(\cdot)$ and let $\{\phi_1^1, \phi_2^1, \ldots, \phi_n^1\}$ be the packet indices under the assumption of hypothesis H_1 using $\phi^1(\cdot)$. Both $\{\phi_1^0, \phi_2^0 - \phi_1^0, \ldots, \phi_n^0 - \phi_{n-1}^0\}$ and $\{\phi_1^1, \phi_2^1 - \phi_1^1, \ldots, \phi_n^1 - \phi_{n-1}^1\}$ should both be independent and exponentially distributed with parameter n_t/n.

Given an exponential distribution with parameter γ, define $pdf(\kappa, \gamma)$ to be the probability density function of this exponential distribution at κ. Therefore, given $\{\tau_1, \tau_2, \ldots, \tau_n\}$ and due to properties of the natural logarithm, if $S_0 = 0$:

$$S_n = S_{n-1} + \ln \frac{pdf(\phi_n^1 - \phi_{n-1}^1, n_t/n)}{pdf(\phi_n^0 - \phi_{n-1}^0, n_t/n)} = S_{n-1} + \ln \frac{pdf(\phi^1(\tau_n) - \phi^1(\tau_{n-1}), n_t/n)}{pdf(\phi^0(\tau_n) - \phi^0(\tau_{n-1}), n_t/n)}.$$

This method of computing S_n is known as the *Cumulative Sum* (CUSUM) method and gives an efficient online method to compute $g(Y^n)$ in the worm detection scenario. The CUSUM method for computing S_n is very efficient in practice if the expression

$$\ln \frac{pdf(\phi^1(\tau_n) - \phi^1(\tau_{n-1}), n_t/n)}{pdf(\phi^0(\tau_n) - \phi^0(\tau_{n-1}), n_t/n)}$$

is generally easily computed. This is because as observations are made in the environment, only a running sum S_n needs to be retained in memory to make continual optimal hypothesis tests instead of full information about past observations. The bottom graph in Figure 9 shows how the CUSUM for S_n changes with time as scanning data is observed for the previously discussed CodeRed1v2 simulation with background radiation.

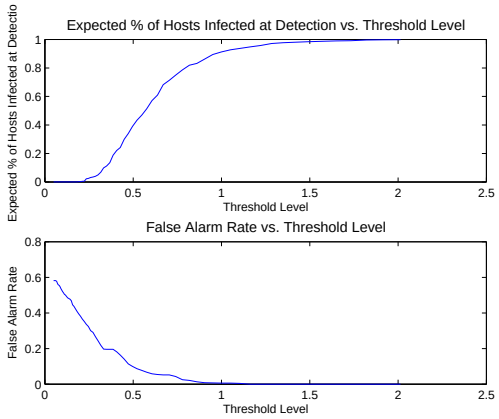

Fig. 10. Plots of the Expected Percentage of Susceptible Hosts which are Infected when a Worm Epidemic is Detected vs. the Log-Likelihood Threshold and the Expected False Alarm Rate vs. the Log-Likelihood Threshold for the CodeRed1v2 SPRT.

8. LIMITATIONS TO DETECTION

Now that the SPRT method for sequential hypothesis testing has been introduced, we discuss next how this optimal detection method can be used to demonstrate fundamental limitations to the detection of RCS worm epidemics even under idealized conditions.

Let us consider a class of SPRT tests for the detection of RCS worms where it is important to decide if a worm epidemic exists very quickly, but it is not important to decide if a worm does *not* exist with great urgency. For the SPRT test in this situation, h should be chosen so that the ARL is relatively small but not so small that the false positive rate α_1 is too high. Similarly, let us assume a lower $-a$ threshold that is very small so that when the SPRT test runs, α_0, the false negative rate, is very small and it is decided that no worm exists very slowly when H_0 holds.

As might be intuitive, there are strong connections between the ARL, the false alarm rates of the SPRT and the threshold levels $-a$ and h. For instance, the ARL monotonically increases with respect to h. This can be seen in first plot of Figure 10 which shows the number of susceptible hosts which are infected when a worm epidemic is detected versus the threshold level h from Matlab simulations for a worm with the CodeRed1v2 simulation parameters discussed above. Similarly, the bottom plot of Figure 10 shows how the false positive rate α_1 depends on the threshold level h for a worm with the CodeRed1v2 parameters discussed above.

Consider the problem of selecting a threshold level h so that a desired ARL is achieved when a worm epidemic is propagating over the Internet. As indicated above, α_1 is monotonically decreasing with respect to h. Therefore, by choosing a desired ARL, this implies there is a fundamental limitation on the false alarm rate α_1 that can be achieved for the given ARL. With the data in Figure 10, it is plotted in Figure 11 how the SPRT false alarm rate α_1 depends on the desired ARL when H_1 holds. Figure 11 indicates that even with full knowledge of the worm and noise parameters, if an RCS worm epidemic should be detected during the early stages of propagation (before 10% of hosts infected), most likely a reasonable false alarm rate (much less than 10%) cannot be achieved for a worm similar

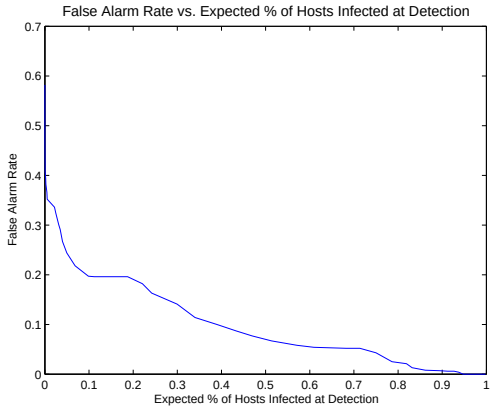

Fig. 11. Plot of the Expected False Alarm Rate vs. the Expected Percentage of Susceptible Hosts which are Infected when a Worm Epidemic is Detected for a CodeRed1v2 worm SPRT.

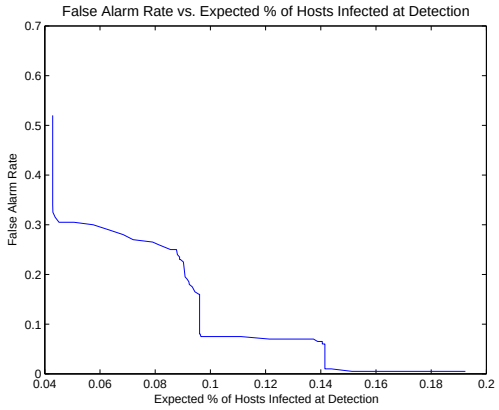

Fig. 12. Plot of the Expected False Alarm Rate vs. the Expected Percentage of Susceptible Hosts which are Infected when a Worm Epidemic is Detected for a Slammer worm SPRT.

to CodeRed1v2, especially when there is no knowledge of the worm parameters. Although hypothesis testing methods have been successful in testing for the existence of infected hosts in a local network [Jung et al. 2004; Weaver et al. 2004], Figure 11 indicates that even optimal hypothesis testing methods under ideal conditions cannot be used to detect worm epidemics by solely observing changes in local port-scanning behavior.

However, as seen in Figure 12, the trade-off between expected detection time and false alarm rate is much more promising for an RCS worm with parameters similar to the Slammer worm. From Figure 12 it can be seen that under the ideal conditions of full knowledge of the worm's parameters, a very small false alarm rate α_1 can be obtained as long as the percentage of hosts infected at detection is at least 14%.

This result at first seems counter intuitive - that a fast worm like Slammer is easier to

detect during its early stages of propagation than a slower worm such as CodeRed1v2. However, when one considers that it is generally easier to detect a faster change in mean packet interarrival time due to a more aggressive worm, this should help to explain why Slammer worms can be detected with better false alarm rates than CodeRed1v2 worms.

9. DISCUSSION

This paper has discussed a number of issues associated with the idealized stochastic properties of RCS worm epidemics. It has introduced density-dependent Markov jump process model for the large-scale propagation behavior of these worms – an approach to worm modeling that has not previously been discussed in the literature. The paper has identified several commonly satisfied conditions under which the variability in the stochastic propagation of RCS worm epidemics predicted by Moore et al. [2003] and Zou et al. [2003] can be ignored. A hybrid deterministic/stochastic model for the observations of a worm's scanning behavior on a local network has also been presented and discussed.

Furthermore, the hybrid deterministic/stochastic worm model has been used to discuss anomaly-based RCS worm detection in the context of detection and estimation theory. An optimal SPRT worm detection method has been proposed under the idealized condition of knowledge of a worm's parameters. Also, fundamental limitations to the detection of RCS worms have been discussed based on simulations of RCS worms with the SPRT detection method. It has been shown that in some sense aggressive RCS worms like Slammer are generally easier to detect than slower RCS worms such as CodeRed1v2.

This paper assumed situations of simple single-vector RCS worms that propagate over hosts with homogeneous connectivity properties. To continue research in this field, methods should be developed to simulate more advanced worms that propagate over non-homogeneous networks. The results in this paper also show that in order to develop effective worm defense strategies, worm detection methods can not rely solely on the detection of port scanning on unused Internet addresses. As part of the ongoing research in worm defense, as new worm detection strategies are developed, there should be a continuing effort to showing the potential weaknesses of these methods, if any exist, in a mathematically rigorous fashion.

REFERENCES

ANDERSSON, H. AND BRITTON, T. 2000. *Stochastic Epidemic Models and Their Statistical Analysis.* Number 151 in Lecture Notes in Statistics. Springer-Verlag, New York.

BASSEVILLE, M. AND NIKIFOROV, I. 1993. *Detection of Abrupt Changes: Theory and Applications.* Prentice-Hall, New York.

DALEY, D. AND GANI, J. 1999. *Epidemic Modelling: An Introduction*. Cambridge University Press, Cambridge.

ETHIER, S. AND KURTZ, T. 1986. *Markov Processes, Characterization and Convergence*. Wiley Series in Probability and Mathematical Statistics. John Wiley and Sons, New York.

GRADSHTEYN, I. AND RYZHIK, I. 1994. *Table of Integrals, Series and Products*, Fifth ed. Academic Press, San Diego. A. Jeffery, editor.

HOEL, P. G., PORT, S. C., AND STONE, C. J. 1971. *Introduction to Probability Theory*. Houghton Mifflin Co., Boston.

JUNG, J., PAXSON, V., BERGER, A. W., AND BALAKRISHNAN, H. 2004. Fast portscan detection using sequential hypothesis testing. In *Proc. of the IEEE Symposium on Security and Privacy*. Oakland,CA.

KERMACK, W. AND MCKENDRICK, A. 1927. A contribution to the mathematical theory of epidemics. *Royal Society of London Proceedings Series A 115*, 700–721.

MODE, C. AND SLEEMAN, C. 2000. *Stochastic Processes in Epidemiology*. World Scientific. World Scientific, Singapore.

MOORE, D., PAXSON, V., SAVAGE, S., SHANNON, C., STANIFORD, S., AND WEAVER, N. 2003. Inside the slammer worm. *IEEE Security and Privacy 1*, 4, 33–39.

MOORE, D., SHANNON, C., AND BROWN, J. 2002. Code-red: A case study on the spread and victims of an Internet worm. In *Proceedings of the Internet Measurement Workshop (IMW)*. Marseille, France.

MOORE, D., SHANNON, C., VOELKER, G., AND SAVAGE, S. 2003. Internet quarantine: Requirements for containing self-propagating code. In *INFOCOM*. San Francisco.

NICOL, D. 2006. The impact of stochastic variance on worm propagation and detection. In *WORM '06: Proceedings of the 2006 ACM Workshop on Rapid Malcode*. Fairfax, VA, USA.

PANG, R., YEGNESWARAN, V., BARFORD, P., PAXSON, V., AND PETERSON, L. 2004. Characteristics of Internet background radiation. In *Proceedings of the 4th ACM SIGCOMM Conference on Internet Measurement*. Taormina, Sicily, Italy.

POOR, H. 1994. *An Introduction to Signal Detection and Estimation*. Springer Texts in Electrical Engineering. Springer-Verlag, New York.

ROHLOFF, K. AND BAŞAR, T. 2005a. The detection of RCS worm epidemics. In *WORM '05: Proceedings of the 2005 ACM Workshop on Rapid Malcode*. Fairfax, VA, USA, 81–86.

ROHLOFF, K. AND BAŞAR, T. 2005b. Stochastic behavior of random constant scanning worms. In *Proceedings of 14th ICCCN*. San Diego, CA, 339–344.

SCHECHTER, S. E., JUNG, J., AND BERGER, A. W. 2004. Fast detection of scanning worm infections. In *Proc. of The Seventh International Symposium on Recent Advances in Intrusion Detection (RAID)*. Sophia Antipolis, France.

STANIFORD, S. 2003. Containment of scanning worms in enterprise networks.

STANIFORD, S., PAXSON, V., AND WEAVER, N. 2002. How to 0wn the Internet in your spare time. In *Proceedings of the 11th USENIX Security Symposium (Security '02)*. San Francisco.

STARK, H. AND WOODS, J. W. 1994. *Probability, Random Processes and Estimation Theory for Engineers*, Second ed. Prentice Hall, Upper Saddle River, NJ.

WALD, A. 1947. *Sequential Analysis*. Dover, New York.

WEAVER, N., STANIFORD, S., AND PAXSON, V. 2004. Very fast containment of scanning worms. In *Proceedings of the 13th USENIX Security Symposium (Security '04)*. San Diego, CA.

WONG, C., WANG, C., SONG, D., BIELSKI, S., AND GRANGER, G. 2004. Dynamic quarantine of Internet worms. In *Proceedings of the International Conference on Dependable Systems and Networks (DSN-2004)*. Florence, Italy.

ZOU, C., GAO, L., GONG, W., AND TOWSLEY, D. 2003. Monitoring and early warning for Internet worms. In *Proceedings of the 10th ACM conference on Computer and communications security*. Washington D.C., USA, 190–199.

ZOU, C., GONG, W., AND TOWSLEY, D. 2003. Worm propagation modeling and analysis under dynamic quarantine defense. In *Proceedings of the 2003 ACM Workshop on Rapid Malcode*. Washington, D.C., 51–60.

ZOU, C., TOWSLEY, D., AND GONG, W. 2004. A firewall network system for worm defense in enterprise networks. Tech. Rep. TR-04-CSE-01, Department of Computer Science and Engineering, University of Massachusetts.

Received Month Year; revised Month Year; accepted Month Year