Infection Quarantining for Wireless Networks Using Power Control

Rahul Potharaju
Department of Computer Science and CERIAS, Purdue University
305 North University Street, West Lafayette, IN 47907 USA
rpothara@purdue.edu

***Abstract*—In recent years, malware has proliferated into wireless LANs as these networks have grown in popularity and prevalence. While the actual effects of malware-related network traffic has been studied extensively in wired networks, analysis has been limited in wireless networks.**

In this work, we investigate a defense strategy based on optimal control that quarantines malware by reducing the communication range of mobile nodes. We characterize how such a solution affects the performance of a wireless network through simulations, leading to a better understanding and prediction of defense protocols that reduce the speed of malware propagation within wireless networks.

***Keywords*-wireless; malware; modeling; implementation;**

I. Introduction

There has been significant research on the topic of malwares with a majority of the research focusing on propagation modeling, detection, and application characterization. Malware spreads through computer networks by searching, attacking, and infecting remote computers automatically. Malware outbreaks such as the Slammer [1] and the Code Red [2] worms in the wired Internet have not only induced expenses in billions but also in a wealth of research [1]–[5].

As the ubiquity of networks continues to grow, there is also an increase in the number of pervasive devices using adhoc communications, allowing direct local interaction between devices in addition to maintaining links to centralized access points. An increaing number of users utilize these devices for a variety of applications such as file-sharing, VoIP etc. Unfortunately, this increased mobility and connectivity creates ground for another propagation vector for spreading malware. Wireless networks differ from wired in the sense that resources are quite limited so a carefully designed malware can cause new forms of havoc. For example, the Zotob worm [6] uses port 445 to infect its victims. Most enterprises block port 445 internally so the worm could have propagated via laptops which were infected while outside the corporate network and subsequently infected other machines after connecting to the intranet. Another example was the Cabir worm [7], which hit the mobile phones in June 2004. Its goal was to drain the battery by excessively using the bluetooth scanning feature present in a mobile phone.

In order to understand the seriousness of future worm threats, there has been significant research into various Internet worm models. In epidemiological research, a number of deterministic and stochastic models have been explored that capture worm spreading dynamics [8, 9]. One of the most popular models is the SIR model developed by Kernack and Kendrick that introduced the notion of compartments which desribed the various phases of the course of an epidemic: *starting* phase depicting slow growth, *explosive* growth and *remission*. It was assumed that the contacts between the members of a population are purely random. This makes it suitable for application to a network of computers including those connected through wireless.

In this work, we investigate a defense strategy based on optimal control [10] that quarantines the malware by reducing the communication range of mobile nodes. The intuition behind the approach is to act in a way that reduces the frequency of contact between the mobile nodes which in turn supresses the spread of the infection. While this decreases the chance of infection, it also decreases the network performance. As it is important for countermeasure design to be able to roughly predict how the performance is affected with different malware models, we propose an agent for ns-2 that models the behavior of the malware as governed by theory. Later, we characterize how a defense strategy such as the one proposed earlier affects performance of a wireless network. This will lead to a better understanding of defense protocols that would reduce the speed of malware propagation within wireless networks.

The paper structure is as follows: We provide an overview of the system model in Section II, the simulation model in Section III, and present initial results in Section IV. We conclude with ongoing and future work in Section V.

II. System Model

We consider the deterministic compartmental SIR model to characterize the worm in our system. In such a deterministic model, individuals in the population are assigned to different subgroups(compartments), each representing a specific stage of the epidemic. The model assumes that the population size in a compartment is differentiable with respect to time and that the epidemic process is deterministic. The SIR model considers a fixed population with three compartments: *susceptible:* $S(t)$, *infected:* $I(t)$ and *recovered:* $R(t)$. At any given time, the following represents the different types of individuals in the system:

- $n_S(t)$ represents the number of individuals not yet infected with the disease at time t(those susceptible to the disease).

- $n_I(t)$ denotes the number of individuals who have been infected with the disease and are capable of spreading the disease to those in the susceptible category.
- $n_R(t)$ represents those individuals who have been infected and then recovered from the disease. Those in this category are not able to be infected again or to transmit the infection to others.

Let the fraction of the infective nodes at time t be denoted by $I(t)$ *i.e.* $I(t) = n_I(t)/N$. Likewise, let $S(t) = n_S(t)/N$ and $R(t) = n_R(t)/N$ represent the fraction of susceptible and recovered nodes at time t respectively. Kurtz *et. al* [11] showed that if N is large, then $S(t)$ and $I(t)$ converge asymptotically to the solution of the following differential equations:

$$\dot{S} = -\beta IS \tag{1}$$

$$\dot{I} = \beta IS - \gamma I \tag{2}$$

$$\dot{R} = \gamma I \tag{3}$$

Here as well as in the rest of the paper, $\dot{X}$ represents the derivative of X with respect to t. Khouzani *et. al* [10] present a containment strategy based on power control that assumes that the reception gain of the susceptible nodes is a variable controlled by the system. Upon detection of malicious behavior, the reception gain of the susceptible nodes can be reduced. This effectively reduces the communication range of the nodes to lessen the frequency of contacts between the infectious and susceptible nodes. This reduces the propagation rate of the infection, thus extending the time available for recovering the infective nodes. Note that the communication range depends on both the transmission and the reception gain of two communicating nodes and reduction of any of these gains reduces the communication range. The set of equations (1), (2), (3) can then be written as follows:

$$\dot{S} = -\beta uIS, S(0) = 1 - I_0 \tag{4}$$

$$\dot{I} = \beta uIS - \gamma I, I(0) = I_0 \tag{5}$$

$$\dot{R} = \gamma I, R(0) = 0 \tag{6}$$

where u is the communication range of the nodes and is the control variable for the system, which is bounded between a maximum and minimum value:

$$u_{min} \leq u \leq 1, u_{min} \geq 0 \tag{7}$$

with the following state constraints:

$$0 \leq S, I, R \tag{8}$$

$$S + I + R = 1 \tag{9}$$

The structural results(variation *w.r.t.* time) of the optimal communication range, u as a function of time, that minimizes the overall system cost which captures desired tradeoffs between communication efficacy (and hence QoS) and containment of the worm was obtained [10] for two cases:

1) Cost function is linear in both u and I

$$J = \int_{t=0}^{T} (CI - u)dt \tag{10}$$

2) Cost function is linear in I but non-linear in u

$$J = \int_{t=0}^{T} (CI + \frac{1}{u})dt \tag{11}$$

where C determines the relative importance (hazard) of the infection. In this work, we consider only the first case and leave the second as part of our future work.

III. Simulation Model

Our simulation consists of N wireless hosts that can reach each other through a routing algorithm (AODV in our implementation). Nodes are assumed to move in a limited region (of area A) and according to the *random waypoint mobility* model. A node stays in one of the three states at any time: *susceptible*, *infectious*, or *recovered*. A node is in "recovered" state when it has been immunized against the infection. This immunity is achieved when it comes into contact with a *security patch* deployed by *healer* nodes. Further, we also assume that *healer* nodes cannot be infected and once nodes have been recovered, they cannot be re-infected. Thus, the state transition of any host can be: *"susceptible-infectious-recovered"* or *"susceptible-infectious"*. When a node becomes infected, it sends out a sequence of infection attempts during its lifetime. At each infection attempt, the infected node scans for neighbors within its communication range to infect and sends out a packet to a random neighbor. If this packet reaches a node, the node becomes infected. The set of neighbors will decrease if the transmission range is lower. Note that the receiver can reduce the sender's transmission range by lowering its reception gain and vice versa. Thus, the transmission range can be considered the control parameter used in minimizing a given cost function. To understand the effects of congestion, we allow susceptible nodes to select a destination (that is not necessarily one-hop away), and transmit packets to it. Such packets need longer to reach the destination as the packets may need to go through multiple-hops. Thus, the delay experienced by legitimate packets increases with the increase of infectious packets. It should be observed that the recovery process is not affected by the reduction in the transmission range (but still depends on the patch distribution frequency or the total number of *healer* nodes in the system along with the state of transmission - whether the *healer* nodes are unicasting or broadcasting security patches).

Our simulation is structured into two separate parts for the sake of simplicity. In the first part, we calculate the switching time t_s, which is defined as the time at which the nodes change their communication range from u_{max} to u_{min} or vice versa (assuming linear dependence shown in Equation 10. Note that for a simulation setting, t_s does not have to exist in which case the communication range remains unaltered. Calculating t_s involves solving the set of differential equations governed by (4), (5), (6) for an optimal t_s to minimize the cost functions (10) and (11) respectively. We achieve this using AMPL [12] equipped with the *snopt* solver [13]. AMPL is an algebraic modeling language for non-linear optimization problems. Later, this switching time is used in a wireless simulation driven by ns-2 [14]. At t_s the reception gain of the nodes is appropriately adjusted. At the beginning of the

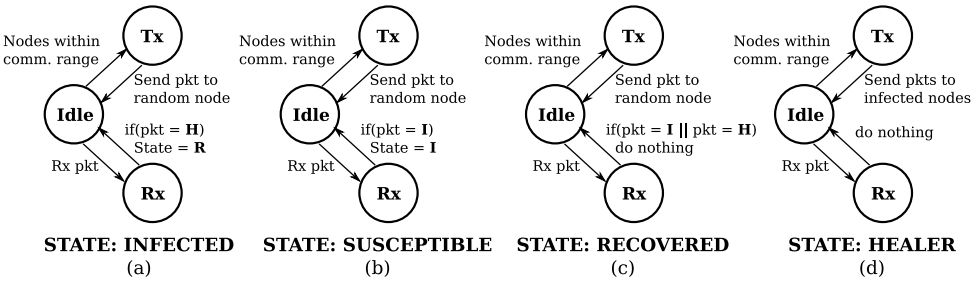


Fig. 1. Behavior of the different types of nodes as modeled in ns-2

wireless simulation, several hosts are initially infectious and the others are all susceptible. Whenever a node comes into the communication range of an infected node, the worm attempts to send out a sequence of infectious attempts. In addition, an infected host will not change its infection behavior if it is infected again by other copies of the worm. Susceptible nodes continue to communicate with other nodes within their communication range. Healer nodes remain unaffected by any range controls applied and continue to patch infected nodes. For the sake of brevity, the behavior of the different types of nodes is depicted in Figure 1.

IV. Results

We first investigate the linear dependence of the cost function on both I and u. Khouzani *et. al* [10] theoretically proved that the optimal communication range for the cost function in (10) is of the *bang-bang* form. As shown in Figure 2, the communication range possesses only two possible values u_{min} and u_{max} and switches abruptly between them at certain *switching times*. We first proceed to verify their result using *AMPL* and *snopt*. The behavior of u is shown in Figure 2. We show the variation in the number of *susceptible, infected* and *recovered* nodes in the system when the communication range is *not* considered. This provides us with basis for behavior we can expect while simulating the scenario using ns-2. From here on, we make it implicit that all experiments assume the parameters shown in Table IV.

Figure 3(a) shows the case where range control is not applied. As the model is deterministic, this behavior represents the baseline for the rest of the experiments. In most cases, the number of infected nodes will either be increasing or decreasing. The key here is to note that there cannot exist a steady state point *i.e.*, the number of infected nodes cannot remain the same as t increases. To see why this is true, consider Equation 5 written as a difference equation:

$$I_{t+1} = I_t + \beta u_t I_t S_t - \gamma I_t = I_t(1 + \beta u_t S_t - \gamma) \qquad (12)$$

If we let $\zeta = 1+\beta u S-\gamma$, this is clearly the epidemic threshold for the model with constant population. The value of ζ decides whether the infection is increasing or decreasing. If ζ is 1, then either $I_t = 0$ or $\zeta_{t+1} < \zeta_t$. If $I_t = 0$ indicates that the infection ended. This case is obvious but if we consider $\zeta_{t+1} < \zeta_t$. ζ depends on three constants $(\beta, \gamma, 1)$ and two state variable(S_t, u_t). If we look at the difference equation(discrete representation) for Equation 4:

$$S_{t+1} = S_t - \beta u_t I_t S_t \qquad (13)$$

we can observe that S_{t+1} decreases whenever $I_t > 0$ and $u_t > 0$. Thus if $I_t > 0$ and $u_t > 0$, ζ_t is decreasing for any $\beta > 0$. This implies that there can never be a steady state where $\zeta_t = 1$ and $I_t > 0$. Further if I_t is increasing, it is doing so at a diminishing rate. And if I_t is decreasing it is doing so at an enhancing rate. Thus in a system with *constant population*, the infection will always disappear in the long run.

Figure 3(b) shows the case where 10% of the nodes are infected but with range control applied. Initially, the communication range is set to u_{min}(normalized to 0 in this case). This prevents the nodes from communicating with each other. Because of this, the system undergoes no change. At $t = t_s = 87$, the communication range is set to u_{max}(normalized to 1 in this case) allowing the nodes start transmitting again. Figure 3(c) shows the system behavior with range control applied to the case where there are 20% infectious nodes. As expected the slope of the infection curve increases after the nodes begin their transmission at the *switching time*.

Figure 4 shows the behavior of the system using ns-2. Figure 4(a) shows the behavior when 10% of the nodes are initially infected and range control is not applied. We believe that the performace of our agent bears close resemblance to the behavior obtained using using *AMPL* and *snopt* and that the minor differences are due to fact that the theoretical model does not consider the network topology and the absolute values of the speed of the nodes. Figure 4(b) and Figure 4(c) shows the scenario with range control applied. The theory assumes that before the *switching time*, nodes can reduce their communication range but still allows the *recovery* process to continue. However, in our model, we do not allow this because when a node reduces its communication range, it will not be able to receive any packets from the *healers*. Also, it may be necessary to bind the cost function to a suitable network metric such as network throughput or node connectivity to get a better picture of the effect of minimizing the cost function. We defer this to our future work.

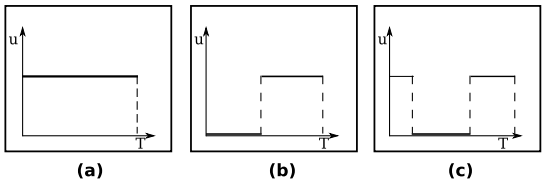
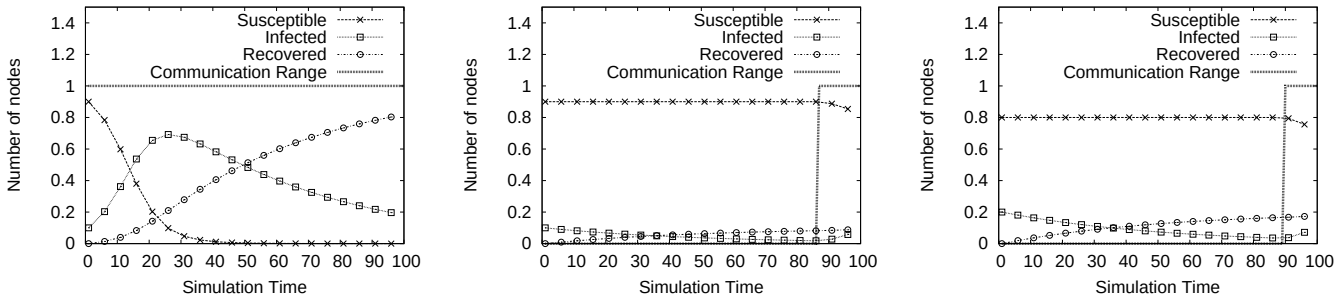
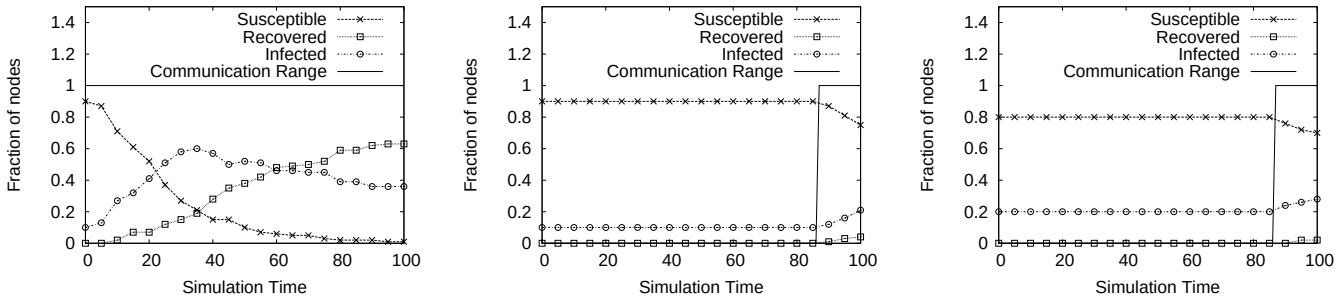


Fig. 2. Bang-bang structure



(a) 10% Infectious Nodes w/o range control (b) 10% Infectious Nodes w range control (c) 20% Infectious Nodes w range control

Fig. 3. **AMPL Simulation Results:** All the figures show the variation of the number of nodes as the simulation time advances. (a) Baseline case where the communication range is unaltered and the system has 10% infectious nodes, (b) The system has 10% infectious nodes and the communication range is altered at t_s to minimize the cost function, (c) The system has 20% infectious nodes and the communication range is altered at t_s to minimize the cost function



(a) 10% Infectious Nodes w/o range control (b) 10% Infectious Nodes w range control (c) 20% Infectious Nodes w range control

Fig. 4. **ns-2 Simulation Results:** All the figures show the variation of the number of nodes as the simulation time advances. (a) Baseline case where the communication range is unaltered and the system has 10% infectious nodes, (b) The system has 10% infectious nodes and the communication range is altered at t_s to minimize the cost function, (c) The system has 20% infectious nodes and the communication range is altered at t_s to minimize the cost function

Parameter	Value
β	0.2
γ	0.02
total	100
infected	variable
susceptible	variable
recovered	0

(a)

Parameter	Value
Power	0.1W
reception gain	variable
healers	variable
frequency	2.4GHz
field area	1000m x 1000m
node speed	20m/s

(b)

TABLE I
PARAMETERS USED IN AMPL AND NS-2 SIMULATIONS

V. FUTURE WORK AND CONCLUSION

We investigate a defense strategy based on optimal control that quarantines malware by reducing the communication range of mobile nodes. We experimentally verify prior theoretical work that proves the structural characteristics of the communication range. We characterize how such a solution behaves in the context of wireless networks through simulations.

Our ongoing work studies the inherent tradeoff between the communication range and packet loss rates. We plan to analyze the effect of different communication patterns on the extent of quarantining and the effectiveness of the mechanisms on data delivery. As a next step, we also intend to propose a suitable mechanism that lets each node independently decide when to reduce its communication range based on certain heuristics.

REFERENCES

[1] D. Moore, V. Paxson, S. Savage, C. Shannon, S. Staniford, and N. Weaver, "Inside the slammer worm," *IEEE Security & Privacy*, vol. 1, no. 4, pp. 33–39, 2003.

[2] C. Zou, W. Gong, and D. Towsley, "Code red worm propagation modeling and analysis," in *Proceedings of CCS*, p. 147, ACM, 2002.

[3] A. Wagner, T. Ubendorfer, B. Plattner, and R. Hiestand, "Experiences with worm propagation simulations," in *Proceedings of the 2003 ACM workshop on Rapid Malcode*, p. 41, ACM, 2003.

[4] J. Kephart and S. White, "Directed-graph epidemiological models of computer viruses," *Computation: the micro and the macro view*, p. 71, 1992.

[5] S. Sellke, N. Shroff, and S. Bagchi, "Modeling and automated containment of worms," in *Proc. of DSN 2005.* , pp. 528–537.

[6] K. Mitnick, "CLICK ME: SOCIAL ENGINEERING IN MALWARE,"

[7] J. Blau, "Cabir worm wriggles into US mobile phones," *PC World*.

[8] R. Anderson and R. May, *Infectious diseases of humans: dynamics and control*. Oxford University Press, USA, 1992.

[9] H. Andersson and T. Britton, *Stochastic epidemic models and their statistical analysis*. Springer Verlag, 2000.

[10] M. Khouzani, E. Altman, and S. Sarkar, "Optimal Quarantining of Wireless Malware Through Power Control," in *Proceedings of the Fourth Symposium on ITA*.

[11] T. Kurtz, "Solutions of ordinary differential equations as limits of pure jump Markov processes," *Journal of Applied Probability*, vol. 7, no. 1, pp. 49–58, 1970.

[12] R. Fourer, D. Gay, and B. Kernighan, "AMPL: A mathematical programming language," *Technical Report, AT&T Bell Laboratories, Murray Hill, USA*, 1987.

[13] P. Gill, W. Murray, and M. Saunders, "SNOPT: An SQP algorithm for large-scale constrained optimization," *SIAM Journal on Optimization*, vol. 12, no. 4, pp. 979–1006, 2002.

[14] S. McCanne, S. Floyd, K. Fall, K. Varadhan, *et al.*, "Network simulator ns-2," 1997.