Cost Optimization in SIS Model of Worm Infection

Jonghyun Kim, Sridhar Radhakrishnan, and Jongsoo Jang

ABSTRACT—Recently, there has been a constant barrage of worms over the Internet. Besides threatening network security, these worms create an enormous economic burden in terms of loss of productivity not only for the victim hosts, but also for other hosts, as these worms create unnecessary network traffic. Further, measures taken to filter these worms at the router level incur additional network delays because of the extra burden placed on the routers. To develop appropriate tools for thwarting the quick spread of worms, researchers are trying to understand the behavior of worm propagation with the aid of epidemiological models. In this study, we present an optimization model that takes into account infection and treatment costs. Using this model we can determine the level of treatment to be applied for a given rate of infection spread.

Keywords—Optimal control, treatment cost, epidemic model, Pontryagin's maximum principle.

I. Introduction

Internet worms have recently become a major problem for the security of Internet networks, causing a considerable amount of resources and time to be wasted on the prevention and treatment of virulent attacks. In general, worms, defined as self-propagating codes, have become much more sophisticated since the Morris worm appeared in 1988 [1]. The task of detection and prevention of worms has become more difficult with our existing computing infrastructure. The convenience of the Internet makes it more vulnerable to malicious Internet exploits. In the area of virus and worm modeling, many studies have employed simple epidemiological models to understand the general characteristics of worm propagation. Epidemiological propagation models have traditionally been used to understand and model the spread of biological infectious diseases [2], [3]. The use of these models for the study of computer worms/viruses has been investigated in [4]-[6]. An analytical model is presented in [5] and [6] that captures the impact of the underlying topology on worm propagation. However, controlling the spread of worms in computer networks has recently become one of the most important issues.

In this letter, we investigate the effect of controlling the spread of computer worms on network delays. A model is presented that can be used to determine the adequate level of control on the total network delay – the delay caused by worms and the delay caused by controlling worm propagation. In addition, we present the optimal control problem of minimizing the total cost of infection which can be interpreted as the network delay incurred by both infection and treatment. This letter focuses on the application of optimal control theory to minimizing the value of treatment costs. We derive the necessary conditions for our cost optimization problem which is solved numerically.

The rest of the letter is organized as follows. Section II presents some of the important issues in computer networks. Section III gives a brief review of the classical epidemic models and a modification of the susceptible-infectious-susceptible (SIS) model with treatment to control Internet worm propagation. The analysis of optimization problems is given in section IV. Our numerical solution of the optimal control problem is also introduced in this section. We conclude the letter with an outline of our future work in section V.

II. Statement of Problems

1. Infection Cost vs. Treatment Cost

This letter considers the optimal control problem of minimizing the value of two costs: first is the infection cost which could impose a constant per period network cost (say, in link bandwidth, network delay, and excessive consumption of

Manuscript received Feb. 20, 2006; revised Aug. 17, 2006.

Jonghyun Kim (phone: +82 42 860 3843, email: jhk@etri.re.kr) and Jongsoo Jang (email: jsjang@etri.re.kr) are with Information Security Research Division, ETRI, Daejeon, Korea.

Sridhar Radhakrishnan (email: sridhar@ou.edu) is with the Department of Computer Science, University of Oklahoma, Norman, USA.

resources) caused by infection; the second is the treatment cost also referred to as network delay incurred by a certain level of filtering countermeasures. In other words, both the infection cost and the treatment cost are considered to be variations of the nodal network delay. Our optimal control approach is proposed, which enables tradeoffs between the infection cost of compromised systems and the treatment cost of defensive countermeasures, with respect to time.

2. Choosing the Number of Nodes to Filter

We present a notion of an optimal number of nodes to receive filtering treatment at a certain infection rate, providing both a mathematical model of the control factors affecting how many nodes to filter, and the numerical solution of our analytical model. For the worm propagation model, we apply the classical SIS model [2], [3]. With this model, the assumptions are a) that during each period the infected nodes can receive treatment that will increase their rate of recovery, and b) they have no preventive properties upon recovery. The treatment will also be assumed to exist in discrete time.

III. SIS Infection Model

Two classical deterministic epidemic models are the basis of our experimental design. In classical epidemic models, a node is defined as an infectious node at time t if it has been infected by a worm before t. A node that is vulnerable to a worm is called a susceptible node.

1. Infection without Treatment

The most common mechanism of infection is through contact with an infected node, and the mechanism of recovery is either deterministic or purely stochastic with a certain typical time of recovery. In the classical SIS model [2], [3], a recovered node immediately becomes susceptible again, while in a more complicated susceptible-infectious-removed (SIR) model [7], cured nodes become immune and effectively excluded from further dynamics. In the SIS model, each node stays in one of two states: susceptible or infectious. Each susceptible node becomes an infectious one at a certain rate. At the same time, infectious nodes are cured and become again susceptible at a different rate. In this model, having the infection and being cured, does not confer immunity. Infectious nodes have a constant probability of recovery in each period with treatment. There is no permanent immunity to the infection, so a cured node becomes susceptible again upon recovery. Using the terms defined in Table 1, the differential equation for the SIS model without treatment is

Table 1. Notations of SIS infection model.

Notation	Definition
N	Size of total vulnerable population
$S(t)$	Number of susceptible nodes at time t
$I(t)$	Number of infectious nodes at time t
$Q(t)$	Number of treated infectious nodes at time t
β	Infection rate
δ	Curing rate of an infectious node
λ	Treatment rate of an infectious node
C_I	Infection cost
$C(Q(t))$	A function of treatment cost
φ	Adjoint variable
ε	Epidemic threshold

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

We assume that at the beginning, $t = 0$, one host is infectious and the other $(N-1)$ nodes are all susceptible. Let $S(t) = N - I(t)$ denote the number of susceptible nodes at time t. Replace $I(t)$ in (1) by $N - S(t)$ and we get

$$\frac{dS(t)}{dt} = -\beta S(t)[N - S(t)] + \delta[N - S(t)]. \tag{2}$$

The solution to (1) is

$$I(t) = \frac{I_0(\beta N - \delta)}{I_0\beta + (\beta N - \delta - I_0\beta)e^{-(\beta N - \delta)t}}. \tag{3}$$

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

$$I_\infty = N - \varepsilon\ , \tag{4}$$

where $\varepsilon = \delta/\beta$ and I_0 is the initial number of infectious nodes. Therefore, not absolutely all of the population gets infected. This shows that each infectious node infects others with an average value of β per unit time.

2. Infection with Treatment

Our optimization model, presented here, takes into account infection and treatment costs. Assuming that filtering treatment is available, infectious nodes can use a level of filtering during each period, which will increase the probability of recovery. The higher the level of filtering, the higher the number of packets which will be processed for treatment and hence the higher the treatment cost. Using (1) and (2), this can be expressed as

$$\frac{dI(t)}{dt} = \beta I(t)S(t) - \delta(I(t) - Q(t)) - \lambda Q(t)\ , \tag{5}$$

where $\lambda > \delta > 0$, and $I(t) \geq Q(t) \geq 0$.

Let $U(t)$ denote the number of untreated infectious nodes at time t. Then (5) can be defined by

$$\frac{dI(t)}{dt} = \beta I(t)S(t) - \delta U(t) - \lambda Q(t), \qquad (6)$$

where $U(t) = I(t) - Q(t)$. If $Q(t) = I(t)$ in each period, then every node which is infected obtains treatment for infection and (6) becomes

$$\frac{dI(t)}{dt} = \beta I(t)S(t) - \lambda Q(t). \qquad (7)$$

Note that if the treatment is very effective, then it may be the case that $I = Q$, and the infection no longer is epidemic with full treatment, which is called the equilibrium state. The equilibrium of the model with full treatment is the same as that of the model without treatment if the recovery from infection is also very effective.

IV. Analysis of Optimization Problems

In this section we consider (time dependent) optimal control strategies associated with infection and treatment cost based on the classical SIS model. To determine the appropriate number of nodes to filter, we need to develop a mathematical model of the potential costs involved in infection and treatment at a given time. We will develop cost functions that system administrators can use to help determine an appropriate level of treatment. Our problem is to minimize the total cost of infection and treatment over the finite time period.

Our objective function to be minimized is

$$\int_0^T [C_I I(t) + C(Q(t))]dt, \qquad (8)$$

where the control function, $Q(t)$, represents the fraction of total infected nodes consuming treatment (to reduce the number of nodes that may be infectious), subject to the infection equation:

$$\frac{\partial I(t)}{\partial t} = \beta I(N-I) - \delta(I-Q) - \lambda Q, \quad I(0) = I_0, \qquad (9)$$

where β, δ, and λ are known positive constants and I_0 is the known initial infected node.

1. Necessary Conditions for Optimization

Our objective function balances the effect of minimizing the number of cases implementing the filtering treatments and minimizing the total cost of infection. The necessary conditions that an optimal control variable must satisfy come from Pontryagin's maximum principle [8], [9]. In order to derive the necessary conditions, we introduce the adjoint variable φ and the Hamiltonian equation, H. This principle converts (8) and (9) into a problem of minimizing a Hamiltonian, H, for the optimization problem:

$$H = C_I I + C(Q) + \varphi[\beta(N-I) - \delta(I-Q) - \lambda Q]. \qquad (10)$$

Furthermore, there exists an adjoint function, $\varphi(t)$, such that

$$\frac{\partial \varphi(t)}{\partial t} = -\frac{\partial H}{\partial I} = -C_I - \varphi(N\beta - 2\beta I - \delta), \qquad (11)$$

where the state problem has initial values $I(0) = I_0$ and the adjoint problem has final values $\varphi(T) = 0$. We then have to minimize H over $0 \leq Q \leq I$, that is,

$$\frac{\partial H}{\partial Q} = C'(Q) + \varphi(\delta - \lambda) = 0. \qquad (12)$$

We denote the marginal cost of treatment as $C'(Q)$, then say $C'(Q) = \alpha Q$ where α is equal to the marginal value of an additional unit of the treatment. Suppose Q^* is an optimal control for the above problem and I^* is the corresponding trajectory so that from (12) the solution for the optimal control is

$$Q^*(t) = \frac{(\lambda - \delta)}{\alpha}\varphi^*(t), \quad 0 \leq t \leq T. \qquad (13)$$

Substituting (13) into (9) gives

$$\dot{I}^* = \beta I(N-I) - \delta I - \frac{(\lambda-\delta)^2}{\alpha}\varphi^*, \quad I^*(0) = I_0 \qquad (14)$$

$$\dot{\varphi}^* = \varphi^*(2\beta I - N\beta + \delta) - C_I, \; \varphi^*(T) = 0. \qquad (15)$$

The optimal control is determined by (13)-(15), that is, we must solve (14) and (15) for optimum trajectory and an adjoint variable. Next, we discuss the numerical solutions of the optimality system and the corresponding optimal control pairs, the parameter choices, and the interpretations.

2. Numerical Results

In this section, we study numerically an optimal treatment strategy for minimizing the total cost of infection since the full dynamic solution of the control problem is usually very difficult and an explicit function-formula does not exist except for very special cases [8], [9]. The optimal treatment strategy is obtained by solving the optimality system, consisting of two differential equations from the state and adjoint equations

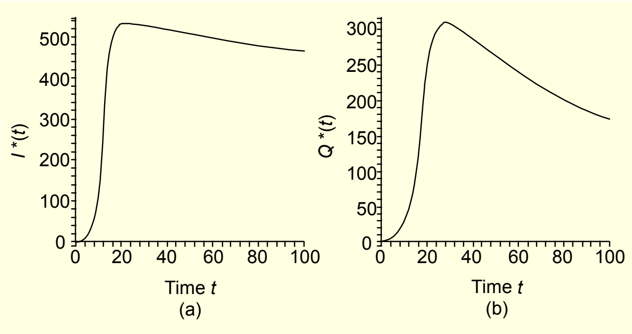

Fig. 1. Optimal control strategy as a function of time.

presented in previous section. Figure 1(a) shows that the average number of infected nodes is plotted as a function of time.

The graph contains 1000 nodes; the infection, cure, and treated rates are $\beta = 1.0$, $\delta = 0.2$, and $\lambda = 0.8$, respectively. In Fig. 1, we assume that the cost weight factor, C_I, associated with the number of infected nodes $I(t)$ is less or equal to the marginal cost of treatment, α, which is associated with a control $Q(t)$. In Fig. 1, the set of the cost weight factors, C_I = 200 and α = 500, is chosen to illustrate the optimal treatment strategy.

Note that, with treatment, the number of infected nodes eventually reaches almost 50% of the total population, and the infection growth slows down after that. However, without treatment, the number of infected nodes reaches almost 80% of the total population [2], [6]. Figure 1(b) shows that for the optimal treatment strategy, the control $Q^*(t)$ is plotted as a function of time with the same parameters as those in Fig. 1(a). This is an expected result because the number of infected nodes consuming a filtering treatment, $Q(t)$, is decreased as the number of infected nodes, $I(t)$, is reduced. In conclusion, the control $Q(t)$ that follows this optimal strategy can effectively reduce the spread of infection and minimize the total cost of infection consisting of both infection and treatment costs.

V. Conclusion

In this letter, we have presented the optimal control problem of minimizing the total cost of Internet worm infection which can be interpreted as the network delay incurred by both infection and treatment. We focus on the application of optimal control theory to minimizing the value of two costs. We have derived the necessary conditions for our optimal control problem which is solved numerically. We have determined that by applying this optimal control strategy, one can very effectively reduce the spread of infection and minimize the total cost of infection.

We are also currently working on the development of effective quarantine techniques, applying our knowledge of cost optimization problems to combating worm infections.

References

[1] E.H. Spafford, "The Internet Worm Incident," *ESEC'89 2nd European Software Engineering Conf.*, Coventry, UK, 1989.

[2] J.O. Kephart and S.R. White, "Directed-Graph Epidemiological Models of Computer Viruses," *Proc. IEEE Symposium on Security and Privacy*, 1991, pp. 343–361.

[3] F. Cohen, "Computer Viruses: Theory and Practice," *Computers & Security*, vol. 6, Feb. 1987, pp. 22–35.

[4] C.C. Zou, W. Gong, and D. Towsley, "Code Red Worm Propagation Modeling and Analysis," *Proc. 9th ACM Conf. on Computer and Communications Security*, Nov. 2002, pp. 138–147.

[5] C. Wang, J.C. Knight, and M.C. Elder, "On Computer Viral Infection and the Effect of Immunization," *Proc. 16th Annual Computer Security Applications Conf.*, 2000, pp. 246–256.

[6] J. Kim, S. Radhakrishnan, and S.K. Dhall, "Measurement and Analysis of Worm Propagation on Internet Network Topology," *Proc. 13th Int'l Conf. Computer Communications and Networks* (*IEEE ICCCN2004*), Oct. 2004, pp. 495-500.

[7] J.C. Frauenthal, *Mathematical Modeling in Epidemiology*, Springer-Verlag, New York, 1980.

[8] G. Knowles, *An Introduction to Applied Optimal Control*, Academic Press, New York, 1981.

[9] M.D. Intriligator, *Mathematical Optimization and Economic Theory*, Prentice-Hall, Englewood Cliffs, N.J., 1971.