International Journal of Soft Computing 5 (2): 56-61, 2010
ISSN: 1816-9503

Final Size Formula for Infected Nodes Due to the Attack of Malicious Agents in a Computer Network

[1]Bimal Kumar Mishra and [2]Prasant Kumar Nayak
[1]Department of Applied Mathematics, Birla Institute of Technology, Mesra, Ranchi 835215, India
[2]Department of Mathematics, G.I.E.T., Gunupur 675022, Orissa, India

Abstract: An attempt has been made to formulate the final size formula for infected nodes in a computer network due to the attack of different malicious agents like viruses, Trojan horse, worms, etc. We assume that the population of the nodes in a computer network is homogenous and there does not exist any heterogeneous mixing. The concept of self-replication of infected nodes and the time lag for self-replication (replication period), latent period and temporary immune period is introduced. The Susceptible Infected Recovered Susceptible (SIRS) class populations is assumed to be bounded by the total size of the population N (t) which is constant at any time instant. The stability of the result is stated in the terms of reproductive number R_0. The system is stable if reproductive number is >1 and unstable if reproductive number is <1. Numerical method is employed to solve the system of integro-differential equations and is used to analyze the behavior of the susceptible, infected and recovered nodes in a computer network.

Key words: SIRS epidemic model, malicious objects, computer network, final size, self-replication, latent period, temporary immunity

INTRODUCTION

Transmission of malicious objects in computer network is epidemic in nature. Malicious object is a code that infects computer systems. There are different kinds of malicious codes such as: Worm, Virus, Trojan horse etc., which differ according to the way they attack computer systems and the malicious actions they perform. Some of them erase hard disks, some others clog the network while some others sneak into the computer systems to steal away confidential and valuable information. Malicious objects can be in any form like attachment of malicious executable file, malicious hyperlink and Phishing.

By clicking incidentally or wrongly an attachment of malicious executable file can infect the system, here the user's awareness is necessary to avoid such type of attacks. If a hyperlink looks likes a spy-ware then by clicking it a user can go to the direction of attack. In a certain sense, the propagation of virtual malicious objects in a system of interacting computers could be compared with a disease transmitted by vectors when dealing with public health.

Concerning diseases transmitted by vectors, one has to take into account that the parasites spend part of its lifetime inhabiting the vector, so that the infection switches back and forth between host and vector (Diekmann and Heesterbeek, 2000). The well-known formula for the final size of an epidemic was published by Kermack and McKendrick (1927) and (Ma and Earn, 2006). They analyzed a simple Susceptible-Infected-Recovered (SIR) model and assumed exponential distribution of infection. Recently, more researchers have stated that the final size formula is valid irrespective of the distribution of infection in a population (Anderson and Britton, 2000; Diekmann and Heesterbeek, 2000; Ma and Earn, 2006).

Whenever, any malicious agent enters a computer network, a matter of immediate interest is the likely magnitude of the outbreak. This is called the expected final size of the epidemic which we denote it as Z in a computer network (Anderson and Watson, 1980; Anderson and May, 1991; Anderson and Britton, 2000; Bailey, 1975; Diekmann and Heesterbeek, 2000).

The formula for Z that Kermack and McKendrick (1927) obtained was totally dependent on the basic reproduction number, R_0 (the expected number of secondary cases caused by a typical primary case in a fully susceptible population) and

$$Z = 1 - e^{-R_0 z}$$

The final size of epidemic in a computer network at any instant t is nothing but the total size of the infected population at that instant t. We arrive to the final size of

Corresponding Author: Bimal Kumar Mishra, Department of Applied Mathematics, Birla Institute of Technology, Mesra, Ranchi 835 215, India

the population in a computer network at any instant of time t by subtracting the susceptible and recovered population from the susceptible population at time zero which is nothing but the total size of the population, N (t). Mishra and Saini (2007a, b) and Mishra and Jha (2007) developed various epidemiological models of transmission of malicious objects in the computer network.

The standard SIRS model equations for the susceptible, infected and recovered classes given by Diekmann and Heesterbeek (2000) is:

$$\frac{dS}{dt} = bS + bR - \mu S - \gamma \frac{SI}{N}$$

$$\frac{dI}{dt} = -\mu I + \gamma \frac{SI}{N} - \alpha I, b > \mu$$

$$\frac{dR}{dt} = -\mu R + p\alpha I$$

Where:

b = Per capita birth rate
μ = Per capita natural death rate
γ = Product of average number of contacts of an individual per unit time and the probability of transmitting the infection during one contact by the infective
p = Probability of temporary immunity acquired when an individual is recovered from the infective class
α = Constant recovery rate

The model developed by Diekmann and Heesterbeek (2000) does not consider time delays like latency period, temporary immunity period, etc., as in the cyber world the recovery is not permanent. We propose the final size formula for infected nodes in a computer network considering the above mentioned time delays and self replication factor and the temporary immunity of the recovered nodes.

Latent period (ω): There is a certain time lag for the node to become infective once it is in the network and is termed as latent period ω.

Temporary immunity period (τ): After the node becomes infected, the malicious object in it may/may not self replicate. Hence, after the run of anti malicious software, the node recovers and attains temporary immunity for a time period termed as period of temporary immunity τ.

Replication factor (r_k): The factor by which any malicious agent self-replicates after infecting any node is called the factor of self replication.

Replication period (ϕ_k): The time lag between a malicious agent infecting a node and its replicated copies becoming infective is called the time for self replication.

Deaths of malicious objects equivalently mean to say the complete recovery of infected files from malicious objects when anti malicious software is run in the computer node for a specific session.

MATERIALS AND METHODS

Mathematical model: We try to find the final size formula for infected nodes in a computer network considering the latency period ω, temporary immunity period τ and time for self replication of kth malicious agent to be constant. The immunity from malicious agents is not permanent but temporary, since in the cyber world, nodes are not permanently immune. We assume that any new node added into the network is susceptible and death rate other then the attack of malicious agents, μ is constant. We further assume that death rate of the nodes due to infection is constant (Deaths of a node equivalently mean to say the isolation of the node which even on running of anti-malicious software may spread malicious agents).

When a node is infected, it may self-replicate with a probability q_k and may not self-replicate with a probability $(1-q_k)$ and when a node is removed from infected class, it may recover with a probability p_k and may not recover with a probability $(1-p_k)$ and that recovery is temporary. Susceptible population is divided into different groups. Nodes may be susceptible due to virus, worms, Trojans, etc. Malicious objects in each group have homogeneous susceptibility but susceptibility of malicious objects from different group is distinct (Mishra and Saini, 2007a, b). Infected population is also divided into different groups (as per their susceptible behavior group). Malicious objects in each group has homogeneous infection but infection of malicious objects from different group is distinct. The flow of malicious objects is shown in Fig. 1. The recovery starts soon after the completion of

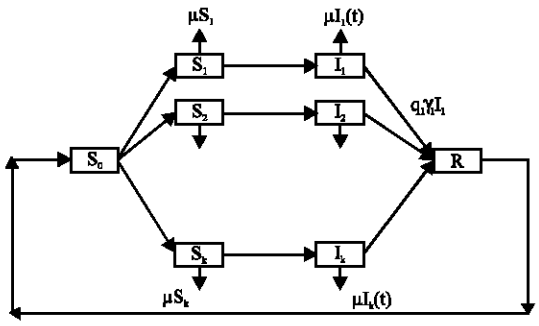

Fig. 1: Flow of malicious agents

the latent period that is the time lag between the start of infection and the start of running of anti-malicious software is considered to be zero. But the rate of infection is assumed to be different from the rate of recovery and so the final size of infected population builds up. As per the assumptions, we get a system of integro-differential equations:

$$\frac{dS_k(t)}{dt} = bS_k + bR_k - \mu S_k - \frac{\gamma_k S_k(t) I_k(t)}{N(t)}$$

$$\frac{dI_k(t)}{dt} = \gamma_k \frac{I(t-\tau)}{N(t-\tau)}.S(t-\tau).e^{-\mu\omega} + \left[q_k \gamma_k \frac{I(t-(\tau+\omega+\phi_k))}{N(t-(\tau+\omega+\phi_k))}.S(t-(\tau+\omega+\phi_k)).r_k.e^{-\mu(\omega+\phi_k)} \right] - (\mu+\alpha_k+\delta)I(t)$$

$$\frac{dR(t)}{dt} = \sum_{j=1}^{n} \left[p_k \alpha_k I_k(t) - \alpha_k I_k (t-\tau)e^{-\mu\tau} - \varepsilon_k R_k(t) \right] - \mu R_k(t) \quad (1)$$

RESULTS AND DISCUSSION

Case I: No birth and no death of nodes: In this case, the birth rate and death rate at any stage are assumed to be zero, i.e., no birth and no death ($b = \mu = 0 = \delta$) which implies that the total population is always constant. The Equations thus obtained from (1) are as follows:

$$\frac{dS_k(t)}{dt} = -\frac{\gamma_k S_k(t) I_k(t)}{N(t)}$$

$$\frac{dI_k(t)}{dt} = \gamma_k \frac{I(t-\tau)}{N(t-\tau)}.S(t-\tau). + \left[q_k \gamma_k \frac{I(t-(\tau+\omega+\phi_k))}{N(t-(\tau+\omega+\phi_k))}.S(t-(\tau+\omega+\phi_k)).r_k \right] - (\alpha_k)I(t)$$

$$\frac{dR(t)}{dt} = \sum_{j=1}^{n} \left[p_k \alpha_k I_k(t) - \alpha_k I_k(t-\tau) \right] \quad (2)$$

We give a formula to find the final size of the infected nodes in a computer network: Final size of infected nodes in a computer network at time t = Total size of the nodes (size of susceptible nodes at time t + size of temporary recovered nodes after the run of anti-malicious software at time t), Thus:

$$Z_k(t) = N(t) - S_k(t) - R_k(t) \quad (3)$$

$$Z_k(t) = N(t) - \int_0^t \frac{dS_k(t)}{dt}dt - \int_0^t \frac{dR_k(t)}{dt}dt$$

$$Z_k(t) = N(t) - \int_0^t -\frac{\gamma_k S_k(t) I_k(t)}{N(t)}dt - \int_0^t \left[p_k \alpha_k I_k(t) - \alpha_k I_k(t-\tau) \right] dt \quad (4)$$

The susceptible and infected population functions are bounded by the total size of population at any time t that is:

$$S_k(t) \le N(t) \Rightarrow \int_0^t S_k(t)dt \le \int_0^t N(t)dt$$

$$I_k(t) \le N(t) \Rightarrow \int_0^t I_k(t)dt \le \int_0^t N(t)dt \quad (5)$$

Now putting these inequalities in Eq. 4, we get the final size formula for the infected population in a computer network at any instant of time t as:

$$Z_k(t) \le N(t)\left[1 + \gamma_k t - p_k \alpha_k t\right] \quad (6)$$

Case II: Birth rate and natural death rates to be positive constants: In this case, we consider the birth rate and natural death rates to be positive constants and both are assumed to be equal, i.e., $b = \mu$ and also assume that recovery is complete and temporary that is there is no death due to infected and no disease induced mortality for recovered nodes i.e., $p_k = 1$. The system of equations thus obtained is as follows:

$$\frac{dS_k(t)}{dt} = bS_k + bR_k - \mu S_k - \frac{\gamma_k S_k(t) I_k(t)}{N(t)}$$

$$\frac{dI_k(t)}{dt} = \gamma_k \frac{I(t-\tau)}{N(t-\tau)}.S(t-\tau).e^{-\mu\omega} + \left[q_k \gamma_k \frac{I(t-(\tau+\omega+\phi_k))}{N(t-(\tau+\omega+\phi_k))}. S(t-(\tau+\omega+\phi_k)).r_k.e^{-\mu(\omega+\phi_k)} \right] - (\mu+\alpha_k)I(t)$$

$$\frac{dR(t)}{dt} = \sum_{j=1}^{n} \left[p_k \alpha_k I_k(t) - \alpha_k I_k(t-\tau)e^{-\mu\tau} \right] - \mu R_k(t) \quad (7)$$

The final size formula for the size of infected population at any time t and putting all the bounded constraints for susceptible and infected class populations as in discussed in Case I gives the following inequality:

$$Z_k(t) \le N(t)\left[1 + \gamma_k t - p_k \alpha_k t\right] \quad (8)$$

From this, it can be easily observed that the inequality obtained in Eq. 6 is identical to Eq. 8. This proves the correctness of theinequality obtained, since the equality

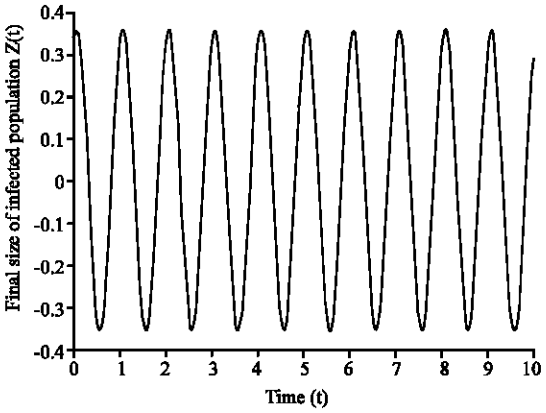

Fig. 2: Variation of final size of infected populaion

of the positive birth and death rates is equivalent to no birth and no death which implies that the total size of the population is invariant of time.

Numerical methods and conclusion: It would be premature to end the discussion by just stating the boundary conditions only. Numerical method is employed to solve Eq. 1-7 under different parameters. Using these in Eq. 8, we observe that the final size of infected population varies as a sinusoidal curve shown in Fig. 2.

In the context of this study, cycle length is the time for one SIRS cycle, i.e., from susceptible stage to completion of temporary immunity after the run of anti-malicious software of the recovered stage.

In a cycle, the size of infected population is initially minimum and it increases gradually and reaches a maximum and as recovery stage starts it gradually decreases. This pattern of variation of the size of infected population repeats in the coming cycles and is periodic in nature.

The susceptible nodes either behave like a cosine curve or exponential curve (Mishra and Saini, 2007a, b). We first assume that the susceptible population varies as a cosine curve as the initial susceptible population for any cycle is maximum and as time passes, the infection increases. This decreases the size of the susceptible population and hence is assumed to vary as a cosine curve. Consider,

$$\begin{aligned} S(t) &= N\cos\left(2\pi(t)\right) \\ R(t) &= A\sin\left(2\pi(t-\omega)\right) + B\sin\left(2\pi(t-\omega-\phi_k)\right) \end{aligned} \tag{9}$$

and using the final size formula Eq. 4, we have

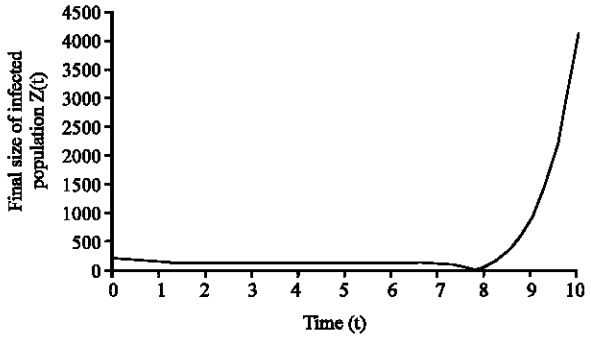

Fig. 3: Variation of final size of infected nodes in the computer network when the infection starts very early in a cycle

$$\begin{aligned} Z_k(t) = N(t) &- \frac{N}{2\pi}\sin\left(2\pi(t)\right) - \frac{A}{2\pi}\cos\left(-2\pi\omega + 2\pi t\right) \\ &- \frac{B}{2\pi}\cos\left(-2\pi\omega + 2\pi\left(-t+\phi_k\right)\right) \end{aligned} \tag{10}$$

The behavior of the susceptible population is also exponential (Mishra and Saini, 2007a, b) as the initial population in the susceptible class in a cycle is maximum and as time increases, the infection increases, the size of susceptible population decreases and reaches a minimum and again with temporary recovery on run of anti-malicious software, the size of susceptible population increases and this carries on periodically for different cycles. The size of susceptible population is assumed to decrease exponentially and the size of infected population increases exponentially. After self-replication with a delay, the size of infected population reaches a maximum in a cycle and it gradually decreases with the recovery stage and this carries on periodically. Consider,

$$\begin{aligned} S(t) &= Ne^{-t} \\ R(t) &= A\left(t-(c-\omega)\right)e^{t-(c-\omega)} + B\left(t-(c-\omega-\phi_k)\right)e^{t-(c-\omega-\phi_k)} \end{aligned} \tag{11}$$

and using the final size formula Eq. 4, we get

$$\begin{aligned} Z_k(t) = N(t) &+ N(t)e^{-t} - A\left[\left(t-(c-\omega)\right)e^{t-(c-\omega)} - e^{t-(c-\omega)}\right] \\ &- B\left[\left(t-(c-\omega-\phi_k)\right)e^{t-(c-\omega-\phi_k)} - e^{t-(c-\omega-\phi_k)}\right] \end{aligned} \tag{12}$$

If the infection starts very early in a cycle, the final size of infected population is constant for most time and after taking a dip it suddenly increases to a maximum value (Fig. 3). If the infection starts very late in a cycle,

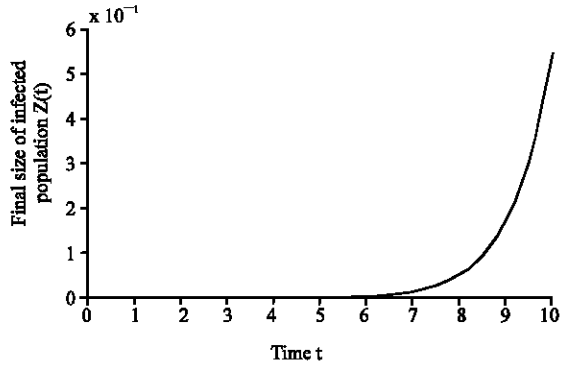

Fig. 4: Variation of final size of infected nodes in the computer network when the infection starts very late in a cycle

the size of infected population is almost zero for most time and after the latent period, it increases exponentially and after the period for self-replication, it drastically increases exponentially to reach maximum and then it gradually decreases with the recovery and reaches a minimum and this repeats periodically (Fig. 4).

Reproductive number: In epidemiology, the basic reproduction number of an infection is the mean number of secondary cases a typical single infected case will cause in a population with no immunity to the infection in the absence of interventions to control the infection. It is often denoted R_0. This metric is useful because it helps to determine whether or not an infectious agent will spread through a population.

When $R_0<1$, the infection will die out with certainty. But if $R_0>1$, there is some possibility of a major epidemic. In particular, the proportion of the population that needs to be immunized to provide immunity and prevent sustained spread of the infection is given by:

$$1-\frac{1}{R_0}$$

We discuss the final size of different nodes under different situations.

Case I : In this case, we consider the birth rate and natural death rates to be positive constants and both are assumed to be equal i.e., $b = \mu$ and also assume that recovery is complete and temporary that is there is no death due to infected and no disease induced mortality for recovered nodes, i.e., $p_k = 1$. For the final size of infected nodes in Eq. 4 that is:

$$Z_k(t) \le N(t)[1+\gamma_k t - p_k\alpha_k t]$$

the proportion of infected population that must be immunized to get temporary immunity is:

$$\frac{Z_k(t)}{N(t)} = 1-\frac{1}{R_0}$$

Substituting the corresponding values, we get:

$$R_0 \le \frac{1}{\gamma_k t - p_k\alpha_k t} \tag{13}$$

Case II: For the final size of infected nodes in Eq. 10 that is:

$$Z_k(t) = N(t) - \frac{N}{2\pi}\sin\left(2\pi(t)\right) - \frac{A}{2\pi}\cos$$
$$\left(-2\pi\omega + 2\pi(t)\right) - \frac{B}{2\pi}\cos\left(-2\pi\omega + 2\pi\left(-t+\phi_k\right)\right)$$

where, A is the recovery rate of the nodes directly recovering from the infected state and B is the recovery rate of the nodes which is infected due to the self-replicated malicious agents, the proportion of infected population that must be immunized to get temporary immunity is:

$$\frac{Z_k(t)}{N(t)} = 1-\frac{1}{R_0}$$

Substituting the corresponding values, we get:

$$R_0 = \frac{N(t)}{\frac{N}{2\pi}\sin\left(2\pi(t)\right) + \frac{A}{2\pi}\cos\left(-2\pi\omega+2\pi t\right) + \frac{B}{2\pi}\cos\left(-2\pi\omega+2\pi\left(-t+\phi_k\right)\right)} \tag{14}$$

Case III: For the final size of infected nodes in Eq. 12 that is:

$$Z_k(t) = N(t) + N(t)e^{-t} - A\left[\left(t-(c-\omega)\right)e^{t-(c-\omega)} - e^{t-(c-\omega)}\right] -$$
$$B\left[\left(t-(c-\omega-\phi_k)\right)e^{t-(c-\omega-\phi_k)} - e^{t-(c-\omega-\phi_k)}\right]$$

the proportion of infected population that must be immunized to get temporary immunity is:

$$\frac{Z_k(t)}{N(t)} = 1-\frac{1}{R_0}$$

Substituting the corresponding values, we get

$$R_0 = \frac{N(t)}{-N(t)e^{-t} + A\left[\left(t-(c-\omega)\right)e^{t-(c-\omega)} - e^{t-(c-\omega)}\right] + B\left[\left(t-(c-\omega-\phi_k)\right)e^{t-(c-\omega-\phi_k)} - e^{t-(c-\omega-\phi_k)}\right]} \quad (15)$$

The final size of infected nodes in a computer network using SIRS epidemic model has been formulated. The boundedness of the final size of infected population has been derived for the system of integro-differential equations. The variation of final size formula for infected population in a computer network under different behaviors of susceptible and recovered populations is analyzed.

CONCLUSION

In this study, the behavior of susceptible population is analogous to cosine and exponential curve whereas the infected population behavior is analogous to sinusoidal curve. The stability of the system is stated in terms of the reproductive number. The basic reproductive rate is affected by several factors including the duration of infectivity of affected nodes, the infectiousness of the malicious agent and the number of susceptible nodes in the computer network. Generally, the larger the value of R_0, the more difficult it is to control the epidemic.

Nomenclature:

S_0 = Inflow population rate
b = Constant birth rate
m_i = Probability of getting susceptible by the ith malicious agent
λ = Infectivity rate
μ = Matural death rate
γ = Infectious rate
δ = Death rate of nodes which are infected due to infection
ε = Disease induced mortality rate for recovered nodes
α = Recovery rate
q_k = Probability of self replication of the kth malicious agent
r_k = Self-replication factor of the kth malicious agent
p_k = Probability of recovery from the attack of the kth malicious agent
$1-p_k$ = Probability of non recovery from the attack of the kth malicious agent
τ = Temporary immunity period
ω = Latency period
ϕ = Time for self replication of kth malicious agent
N = S + I + R, the total population size

REFERENCES

Anderson, D. and R. Watson, 1980. On the spread of a disease with gamma distributed latent and infectious periods. Biometrika, 67: 191-198.

Anderson, H. and Britton, 2000. Stochastic Epidemic Models and their Statistical Analysis. Vol. 151. Springer-Verlag, New York.

Anderson, R. and R. May, 1991. Infectious Diseases of Humans: Dynamics and Control. Oxford University Press, Oxford.

Bailey, N.T.J., 1975. The Mathematical Theory of Infectious Diseases and Its Application. 2nd Edn., Griffin, London.

Diekmann, O. and J.A.P. Heesterbeek, 2000. Mathematical Epidemiology of Infectious Diseases: Model Building, Analysis and Interpretation. John Wiley and Sons Ltd., New York.

Kermack, W.O. and A.G. McKendrick, 1927. A contribution to the mathematical theory of epidemics. Proc. R. Soc. A, 115: 700-721.

Ma, J. and D.J.D. Earn, 2006. Generality of the final size formula for an epidemic of a newly invading infectious disease. Bull. Mathematical Biol., 68: 679-702.

Mishra, B.K. and D.K. Saini, 2007a. Mathematical models on computer viruses. Applied Math. Comput., 187: 929-936.

Mishra, B.K. and D.K. Saini, 2007b. SEIRS epidemic model with delay for transmission of malicious objects in computer network. Applied Math. Comput., 188: 1476-1482.

Mishra, B.K. and N. Jha, 2007. Fixed period of temporary immunity after run of anti-malicious software on computer nodes. Applied Math. Comput., 190: 1207-1212.