AGENT-BASED MODELING OF VIRUS PROPAGATION IN A NETWORK

Myelinda A. Baldelovar, MIT
Surigao del Sur State University-Tagbina Campus
Tagbina, Surigao del Sur
Philippines
myeandante@yahoo.com

Abstract

Computer virus is a computer malware that when executed, replicates by inserting copies of itself into other computer programs, data files, or the boot sector of the hard drive. The study utilizes an experimental design using NetLogo simulation technique. Simulation is designed to determine the virus propagation of computers connected in a computer network. This study determines the effect of virus scan frequency and rate of recovery of computers connected and the impact of their interaction in a network. Results reveal that only virus scan frequency has a significant impact in controlling the virus propagation in a network. This study is important to understand virus propagation behavior and future threats in order to develop mechanisms and policies to prevent it.

Keywords: Netlogo, program, simulation, virus scan frequency, virus

Introduction

One of the vital aspects of computer security is computer viruses. Computer viruses are programs that can infect other programs by modifying them to include and possibly evolved, copy of it. With the infection property, a virus can spread throughout a computer system or network using the authorizations of every user using it to infect their programs. This study aims to determine the following: a) impact of virus scan frequency and recovery rate of computers in the spread of virus in a network and b) the interaction between virus scan frequency and recovery rate of computers in a network.

The spread of computer viruses is a non-linear dynamic system which is similar to the spread of epidemics in human populations (Kephart,1993). Daley et. al, (2005) states that modeling of infectious diseases is a tool which has been used to study the mechanisms by which diseases spread, to predict the future course of an outbreak and to evaluate strategies to control an epidemic . Kermack et. al, (1927) created the SIR model in which a fixed population with three compartments was considered. These compartments are susceptible (S), infected (I), and removed (R). This model is used in epidemiology to compute the amount of susceptible, infected and recovered people in the population.

Previous studies showed spread of an epidemic can be predicted if initial number of susceptible people (S(0)), the infection rate (K), and the removal rate (by quarantine or cure) (Q).Many studies were done in the past about virus propagation in the network. However, the impact of virus scan frequency, and rate of recovery in the spread of virus in the network was not explored well.

This study simulates the virus propagation in a computer network to predict its occurrence. Predicting the virus spread in a local area network is essential to forecast its behavior and outline countermeasures to prevent its occurrence. Through this, an understanding of the role of the virus scan frequency and recovery rate on the spread of the disease will be developed and policies will be formulated.

Model Definition

The theory of the study Epidemic Propagation Models (Bailey, 1975) is applied in modelling the spread of viruses in computers. Among these models is the SIR (Susceptible-Infected-Recovered). S-I-R model is the standard model for the analysis of computer virus infections. There are three states in each node in a network. The node can be susceptible, infected, or recovered.

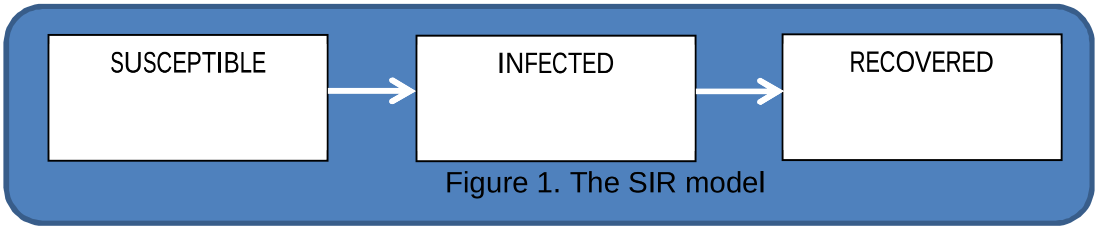

Figure 1. The SIR model

- A SUSCEPTIBLE node is infected but susceptible to virus and can be infected by its neighbors.
- An INFECTED node can infect its neighbor according to infection probability.
- A RECOVERED node are immune to the infection and do not affect the transmission dynamics in anyway when they contact individuals.

Movement of individuals is one-way only, $S \to I \to R$ and the rate constant that control how fast members progress into the *Infected (I)* and *Recovered (R)* groups respectively are the two fundamental parameters of the model which are a) infection rate and b) recovery rate. A composite parameter is often used and is referred to as the contact number.

Research Design and Methods

This study will utilize the existing Virus on a Network model of Wilensky (2008) found in the net logo models library with the changes in the parameter definition, which is shown in Table 1. To determine the percentage of infected computers, the scheduled virus scan and rate of recovery are controlled during simulation. Each parameter is set at two levels, high (1) and low (0), hence, the experimental design is a 2^2-factorial experiment design. This experimental design yields four (4) possible treatment combinations and every combination will generate ten (10) observations.

Table 1. Analysis on the Parallelism of Parameters Used in Different Model

Parameters in Virus on a Model	**Parameters in Virus Spread Model**
Number of Nodes	Number of Computers
Average Node Degree	Average number of connected computers
Initial Outbreak Size	Initial Infection Size
Virus Spread Chance	Virus infection rate
Virus Check Frequency	Scheduled virus scan
Recovery Chance	Rate of Recovery
Gain Resistance Chance	Rate of Resistance

The following are the parameters used in order to determine the behavior of virus propagation in a local area network.

1. Number of computers is 300
2. Average number of computers connected in a computer is 6
3. Initial infection size is 3
4. Scheduled virus scan
5. The number of infected computers depends upon the Virus rate of infection, Rate of Recovery and Rate of Resistance.

Average number of computers connected in a network – the average number of links coming in of each computer

Initial Infection size – number of infected computers at the start of simulation

Scheduled virus scan – a virus scan whether it is infected by a virus or not.

Virus infection rate – the probability that susceptible nodes will be infected.

Rate of recovery – the probability of virus removal after it has been detected

Rate of resistance – the probability that the node will become resistant to the virus after recovery

Results and Discussion

Table 2. Analysis of Variance Table

Sources	***Sum of Squares***	***Degrees of Freedom***	***Mean Square***	***F-value***	***p-value***
T1 (Virus scan frequency)	5123.4	1	5123.4	165.73	0.000
T2 (Rate of recovery)	11.6	1	11.6	0.37	0.545
Interaction	6.0	1	6.0	0.19	0.662
Error	1112.9	36	30.9		
Total	6253.9	39			

Data in Table 2 above is the analysis of variance result after performing simulation trials. It reveals that virus scan frequency is significant and has a positive impact in controlling the spread of virus in a network which means that more frequent virus scan can reduce the chance of virus spread. On the other hand, the rate of recovery has no significant effect in controlling the spread of virus. However, the interaction between virus scan frequency and rate of recovery is insignificant and this signifies that it has no impact in controlling the spread of virus in a computer network.

V. CONCLUSION

Virus on a Network model of Wilensky (2008) found in the NetLogo models library was used in this study in order to determine the virus behavior in a computer network. The virus scan frequency and computer rate of recovery values were controlled during simulation to determine its impact on the spread of virus in a network. It can be noted from the results that the frequency of virus scan can reduce the chance of virus spread in a network.

REFERENCES

1. Cohen, F.: Computer Viruses. PhD thesis, University of Southern California (1985)
2. Chen, Li-Chiou, The Impact of Network Topology on the Spread of Anti-Virus Countermeasures , 2003 http://www.casos.cs.cmu.edu/publications/papers/chen_2003_impactnetwork.pdf
3. J. Ren, X. Yang, Q. Zhu, L.-X. Yang, and C. Zhang, “A novel computer virus model and its dynamics,” Nonlinear Analysis: Real World Applications, vol. 13, no. 1, pp. 376–384, 2012.
4. M. Draief _, A. Ganesh y and L. Massouli_e , Thresholds for virus spread on networks zhttp://www.iis.ee.ic.ac.uk/~m.draief/file/Home_files/Thresholds%20for%20virus%20spread%20on%20networks.pdf
5. Stonedahl, F. and Wilensky, U. (2008). NetLogo Virus on a Network model. http://ccl.northwestern.edu/netlogo/models/VirusonaNetwork. Center for Connected Learning and Computer-Based Modeling, Northwestern University, Evanston, IL.
6. Wilensky, U. (1999). NetLogo. http://ccl.northwestern.edu/netlogo/. Center for Connected Learning and Computer-Based Modeling, Northwestern University, Evanston, IL.
7. http://www.ukessays.com/essays/information-systems/about-computer-viruses.php
8. *Kermack, W. O.; McKendrick, A. G. (1927). "A Contribution to the Mathematical Theory of Epidemics". Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* ***115*** *(772)*
9. Daley, D. J. & Gani, J. (2005). *Epidemic Modeling: An Introduction.* NY: Cambridge University Press.