Indonesian Journal of Electrical Engineering and Computer Science
Vol. 23, No. 1, July 2021, pp. 370~377
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v23.i1.pp370-377

Modeling virus spread on a network using NetLogo for optimum network management

Catherine R. Alimboyong
Department of Computer Studies, Surigao Del Sur State University, Tandag City, Philippines

Article Info

Article history:

Received Oct 1, 2018
Revised Dec 10, 2018
Accepted Jan 25, 2019

Keywords:

Computer virus
NetLogo
Network security
SIR

ABSTRACT

The infections in computer networks are complex. Its spread is analogous to a contagious disease which can cause destruction within a few seconds. Viruses in a computer or computer networks can spread rapidly by various means such as access to online social networking sites like twitter, Facebook, and opening of email attachments. Thus, infections can go from being little dangerous to significantly harmful for a network. This paper proposed a simulation model that can predict the propagation of virus including the trend and the average infection rate using NetLogo software. Observed and simulated data sets were validated using chi-square tests. Results of the experiment have demonstrated accurate performance of the proposed model. The model could be very helpful for network administrators in mitigating the virus propagation and obstruct the spread of computer virus other than the usual prevention scheme particularly the use of antivirus software and inclusion of firewall security.

Corresponding Author:

Catherine R. Alimboyong
Department of Computer Studies
Surigao Del Sur State University
Rosario, Tandag City, Surigao del Sur, Philippines
Email: cralimboyong@sdssu.edu.ph

1. INTRODUCTION

Malware-infested computer systems and networks are a common phenomenon in the age of big data. Malware is a broad term that encompasses computer viruses, worms, rootkits, trojan horses, adware, spyware, and other types of unwanted software that aim to degrade system performance [1], [2]. Recently, malware is being developed more and more to affect computer networks and spread over networks to damage the computer as much as possible [3]. Therefore, a malware-infested computer implies that the entire network may be affected in a short time, extending downtime [4]. Each minute of downtime means loss of income and system performance. In most circumstances, this could result in severe crises like concens about security. Therefore, it is very important to understand how malware spreads and affects the networks [5], [6]. In the era of information and communication technology (ICT) systems, everyone thinks as to how life is going to be carried out as ICT and communication devices play a crucial role in the different fields of our daily life particularly in this time of pandemic crisis. Due to the fact that the attack of virus on networks is harmful and that it can damage hardware or software, the research on network virus has become vital [7]. This study attempts to generate a model of how virus spreads on a network using a NetLogo software v.5.3.1. in order to improve the strategies of controlling the infestation among networks.

Malware attacks on computer networks are becoming more common, and researchers are working to better train network managers to avoid hugely catastrophic threats. Dubey *et al*. [8] affirmed that email Worms not only decrease efficiency, resulting in a loss of time and resources, but they also have an effect on

intangible assets such as a company's reputation and customer loyalty [9]. Infections can thus progress from being relatively harmless to being extremely harmful to a network [10]. Several studies have reported attacks from internet sources such as incidents of distributed denial of service (DDoS), attacks by large distributed organizations like Yahoo, Amazon, CNN.com and other webpages in 2000. According to Arbor Networks' 2010 survey, there was a startling 102 percent increase in DDoS attacks in 2010 compared to 2009 [6], [11], [12]. At present, there are numerous mechanisms developed governing the spread of malware around the networks such as mathematical models capable of accurately foreseeing the propagation of epidemic malicious software. The SIR model of Maji *et al*. [4]; Upadhyay *et al*. [7] where a fixed population of necessary aspects was taken into account. These features are susceptible (S), infected (I), and removed (R). In epidemiology, this method was used to evaluate the number of susceptible, sick, and recovered people in a population [13]. Deepened by the study of Zhang *et al*. [14], simulating of infectious diseases is a technique that has been used to research the processes as to how infections are transmitted, forecast an outbreak's potential path, and test approaches for managing an outbreak, according to the authors.

Conversely, computers with installed software are not indefinitely virus resistant, but they are checked for the virus on a regular basis, and the software removes it if the device is found to be infected. A device may be infected by the same virus several times and stay infected until the antivirus software runs another scan [6], [15]. Alternatively,when viruses mutate, the contact process also roughly describes the spread of epidemics in the presence of regularly updated antivirus software, which confers permanent immunity. In this case, antivirus software stops a device from being infected with the same virus, but not with all mutated variants [16]. Several studies have been conducted and many mathematical models and strategies have been proposed to at least obstruct the spread the virus on networks, SIR model [4], SIRS model [13], SEIR model [17], SEIQR model [18], SIPS model [19], SAIR model [20], [21], SLBRS [22], SIIRS model [23], delayed model [24] on virus propagation.

In this work, a new model describing the spread of viruses on a network is introduced as there are still some deficiencies and issues in model security analysis and control strategies [7]. An attempt has been made which is very useful on the part of network administrators in conveying how systems could be configured to prevent or, at the very least, slow the spread of malware to the greatest extent possible. Using the model, it present three states of a node in a network (a) susceptible (b) infected and (c) resistant and how they rank in terms of malware propagation vulnerability. This study would be extremely beneficial to organizations in making network infrastructure design decisions and can assist them in developing or formulating policies for optimum network configuration. The following is how the rest of the article is organized: The model definition and formulation are given in Section 2, the results and discussion is presented in Section 3, and the conclusion is presented in Section 4.

2. MODEL DEFINITION AND FORMULATION

This study utilizes the simulation method of research. This technique is used to simulate the spread of computer virus on networks. The main focus of this study is to generate a model by visualizing how the system simulates the infestation of virus on networks in relation to the parameters pointed out in this paper. It is deemed necessary to understand the mechanism on how these viruses spread so that a possible control can be designed; in this end, security and safety of files and important documents can be ensured. Chi-square tests were used to examine if any significant differences existed between the observed and simulated data

2.1. Model definition

This paper endeavors to model how a virus spreads across networks using the Virus on Networks from the models library NetLogo v.5.3.1 (version 2016) [25]. Uri Wilensky directed the creation of an agent-based software package called NetLogo at Northwestern University's Center for Connected Learning and computer-based modeling (CCL). It is the most notable case of a multi-agent simulator that includes StarLogo, which Wilensky and Mitchel Resnick designed at the MIT Media Lab [26].

NetLogo shows what can actually take place when turtle populations are given series of requirements to follow. Given its user-friendly programming interface, NetLogo can handle complex simulation as well as the ability for advanced programmers to add their own Java extensions. As a consequence, NetLogo is utilized by diverse group of people, ranging from elementary school students to academics in the social, electronic, and hard fields of science. On the NetLogo website, a download area, template pages, sample downloadable extensions, user guides, a FAQ, and links to various resources are all available [26].

Since there are observations which indicate that computer even if installed with antivirus software are still found to be infected after being scanned [6]. The same virus can infect multiple machines. several times, and it remains infected each time until the antivirus program runs another scan [1], [6]. Since computer

viruses and epidemic diseases have similar propagation characteristics, some researchers have proposed mathematical models to represent the spread of computer worms over networks in recent decade [6], [27], [28].
By looking at the propagation, the researcher considers this model which relies on the following assumptions:

- Majority are using computers to check emails, so chances of virus infestation are expected.
- People use and interact with networks through online social networking sites like twitter, facebook. in which the average of network usage can be drawn-this leads to the spread of virus.
- On any given instance, an email or file containing virus is accessed, it becomes active; therefore, a computer or node becomes active-resistance can be measured or drawn.
- Computers without scanning or protection through antivirus software suffer maximum damage or shut down because of the attack.

Figure 1 shows the susceptible-infected-recovered (SIR) model. The S-I-R model is the most widely used method for analyzing computer virus infections (Maji *et al*., 2020). Each node in a network may either be in one of three states. That is, the node can be susceptible, infected, or recovered. The model simulates the spread of virus on networks. The variables are utilized and processed in a Net Logo software to outline the average of infected nodes. The model and the program code adopted from Wilensky [25] are shown in the following section below. Individual movement is one-way **S→I→R** and the rate which control how quickly members progress into the Infected (I) and Recovered (R) groups are the model's two fundamental parameters, namely a) infection rate and b) recovery rate. A composite attribute, classified as the contact number, is frequently used.

Figure 1. SIR model [4]

2.2. Parameters

The propagation of a virus is determined by all of the parameters set. The following criteria are considered when determining how computer virus spreads on networks: (a) initial number of nodes; (b) average e-mail check/file download; (c) average network usage/access; and (d) frequency of virus scanning. Due to the complexity of the mail network and the uncertainty of the behavior of email users, this paper relies mainly on simulation rather than mathematical analysis. It is in this way, a realistic scenario for the spread of the virus is presented.

The parameters shown in Table 1 are taken from Wilensky (2016)'s Virus on a Network model [26] in the models library. In order to determine the behavior of spread of viruses on networks, the following parameters are defined.

- Initial number of nodes-number of nodes at the start of simulation.
- Average of email check/file download-the average number of email check/file download in each node.
- Average of network usage-the average number of network usage in each node.
- Frequency of virus check-the number of instances a virus scan is performed for each infected node.

Table 1. Consistency of parameters to be defined for computers in a network

Parameters on VIRUS	Parameters of Virus on networks
Initial-nodes	Initial-nodes
Average of infectious nodes	Average of infectious nodes
Average of immune nodes	Average of immune nodes
Average of Virus Check Frequenncy (0.00 times/year)	Average of Virus Check Frequenncy (0.00 times/year)

2.3. Scenarios from the model

Table 2 shows the parameters and values that are used to determine the action of virus propagation in a network at the start of the simulation. By looking at Figure 2, one is able to visualize how the system simulates the behavior of virus propagation on networks in relation to the average of email check/file download, average of network usage, and frequency of virus check (scanning). The simulation scenario

presented in Figure 2 is sufficiently accurate to permit network administrators develop and formulate a comprehensive policy to reduce if not impedes the spread of virus.

Table 2. Parameters of virus on networks actual data

Parameters of virus on networks	Value
Initial Number of nodes	85
Average of email check/file download	65
Average of network usage/access	70
Frequency of virus check (scanning)	50%
Average Number of infected nodes	16

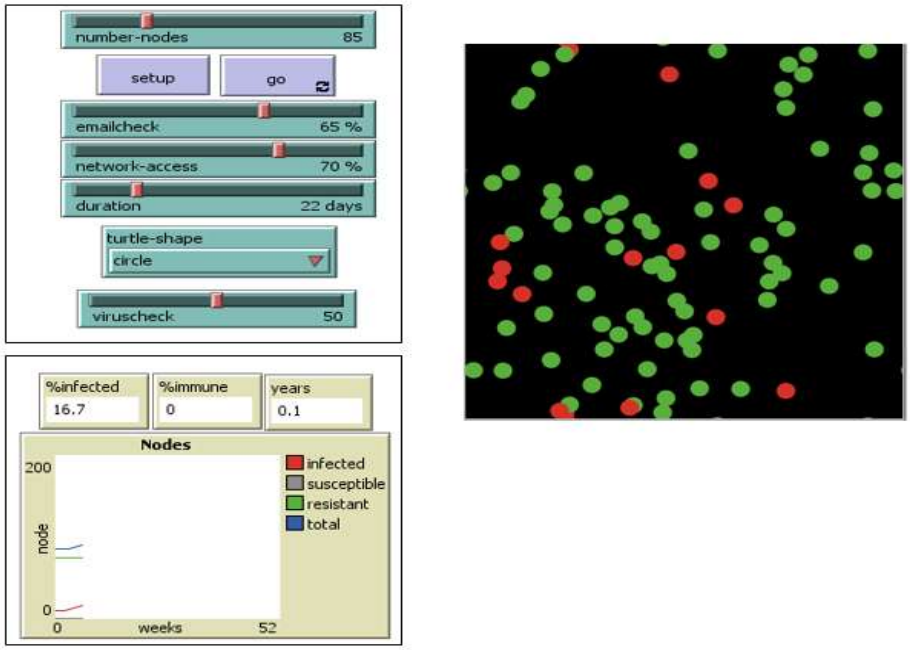

Figure 2. Scenarios from the model (NetLogo software)

2.4. Mathematical analysis

In this model, a chi-square goodness of fit test decides whether or not a sample of data is representative of the population. The goodness of fit test is used to make sure whether sample data is representative of the general population (i.e. a population with a normal distribution or one with a Weibull distribution). In other words, it informs you whether the data in your sample is indicative of the data contained in the entire population [29]. Equation 1 shows the chi-square formula used in this analysis.

$$\chi_c^2 = \sum \frac{(O_i - E_i)^2}{E_i} \quad (1)$$

where:
O_i = the observed frequency (the observed counts in the cells)
E_i = the expected frequency if NO relationship existed between the variables

2.4. Simulation

In this section, the results of simulations are presented in order to better understand how virus spread. Table 3 and Figure 3 presents the actual data collected from the university's network administrator. Data set of infected nodes from January to December were collected from the report presented by the network administrator where N=85 are the number of available computers (nodes) utilized by the students in the internet laboratory. The study considers ten (10) observations as can be seen in Table 3. Using the mean, the total number of infected nodes over a 12-month span is 15.83.

Table 3. Actual data gathered from 85 nodes

Parameters (N=85)	Trial 1 (Jan)	Trial 2 (Feb)	Trial 3 (Mar)	Trial 4 (Apr)	Trial 5 (May)	Trial 6 (June)	Trial 7 (Jul)	Trial 8 (Aug)	Trial 9 (Sept)	Trial 10 (Oct)	Trial 11 (Nov)	Trial 12 (Dec)	Average of infected nodes
Infected	18	16	19	14	12	15	15	18	16	18	15	14	15.83

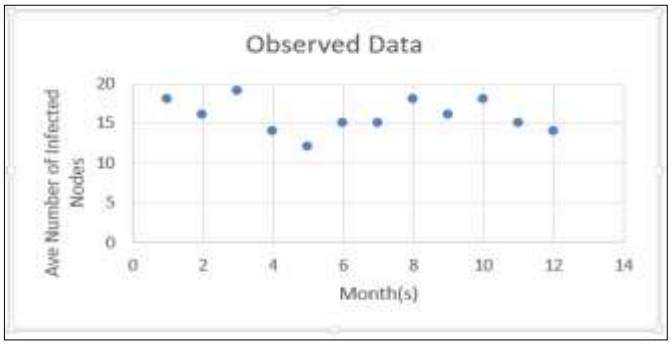

Figure 3. Actual/observed data

Furthermore, the graph in Figure 3 depicts the data gathered from the university's network administrator, specifically in Surigao del Sur State University, Tandag City, Philippines. The strength of virus transmission is best described by the plot in the graph, which is calculated by the probability of opening of email attachments and network use. It also depends on the amount of time and frequency at which viruses are scanned. The result is affirming to the claim of M. Zhang *et al*. [18] that large-scale and rapid virus dissemination occurs during the most frequent access to emails. As Upadhyay *et al*. [5] puts it, if susceptible computers come into contact with infected computers, the virus can spread swiftly

3. RESULTS AND DISCUSSIONS

Computer virus has become one of the major threats to the security of the network. It has rapidly evolved across the internet, causing millions or even billions of data loss [29]. At present, access to online social networking sites (such as twitter, facebook.) that eventually attracted people of all ages from around the globe is the primary vehicle for transmitting computer viruses [4]. Computer viruses, on the other hand, are typically delivered to a computer as an attachment to an email message, which when activated by the user sends extended copies of the virus to other recipients. Based on the actual data sets, the researcher attempts to establish a simulation environment which includes 85 nodes. Then ten (10) simulation runs were performed given the initial data collected.

Table 4 and Figure 4 presents the simulation of ten observations or trials. Hence, the average number of infected nodes which is 15.03 over a period of 12 months, closely exhibits and has almost the same value from the average of infected nodes of the actual data which is 15.83. The result is evident after ten (10) simulation runs has been performed. This implies that the simulation model could reflect the behavior of how a virus spread over a network. Similarly, it can help predict the propagation of virus including the trend which helps network administrators prevent and control the virus spread.

Table 4. Simulated data

Parameters (N=85)	Trial 1 (Jan)	Trial 2 (Feb)	Trial 3 (Mar)	Trial 4 (Apr)	Trial 5 (May)	Trial 6 (June)	Trial 7 (Jul)	Trial 8 (Aug)	Trial 9 (Sept)	Trial 10 (Oct)	Trial 11 (Nov)	Trial 12 (Dec)	Average of infected nodes
Infected	16.47	16.67	12.94	14.12	12.96	11.76	19.05	16.47	15.29	14.12	17.65	12.94	15.03

Figure 5 depicts the simulation result of both scenarios. The figure revealed that the actual and simulated data is very close. Thus, it implies that the model after being validated, is sufficiently accurate and

consistent. In terms of implementation, however, there is little that network administrators can do to control the spread of a virus in the world at large. The result show consistency with the study of Tang *et al*. [7], as he puts it that in the process of virus prevention and control, we must not only improve people's understanding of virus transmission mechanisms, but also evaluate the security of a complex environment using the safety entropy of network networks and their evolving patterns.

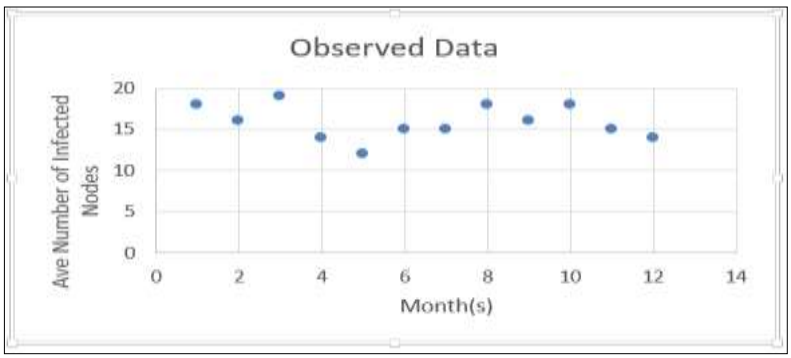

Figure 4. Simulated data

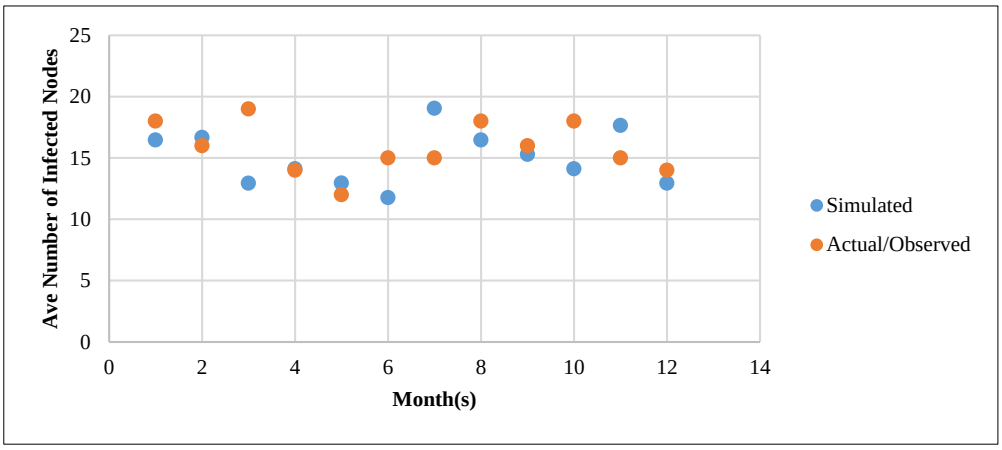

Figure 5. Simulation of both scenarios (simulated and observed)

Consequently, whilst the efforts of network administrators are usually focused on minimizing damage once the infection enters the computer system for which they have responsibility. This findings suggest that it is vitally important to impose not only to network administrators but also to the internet users with network security awareness trainings and to be able to take proper measures to renovate the system, making it strongly-protected [4], [30]. Furthermore, the result clearly shows that the spread of virus can be predicted using the proposed model. Thus, the model enhances and certainly expands the process of scanning the computers as a way of addressing the issue of network security. The increasing number of infected nodes continues to challenge the network administrator to upgrade the software periodically and to increase security awareness. At multiple test runs, the simulation model can predict the propagation of virus including the trend, which is of big help for network administrators in obtaining optimum network management.

4. CONCLUSION

The simulation showed that spread of virus on networks can be minimized and predicted. To simulate virus spread, Wilensky's Virus on a Network model was used in a computer network and data sets were validated using chi square test. Results of the experiment have demonstrated accurate performance of the proposed model. The study suggests that scanning can be performed weekly or perhaps regularly at a specified time frame. Moreover, the simulation experiment made in this study significantly improves

epidemic eradication as compared to employing a lone strategy like the current virus prevention scheme utilizing antivirus software and firewall security.

ACKNOWLEDGEMENTS

The author expresses gratitude to the evaluators for their insightful comments. This work is supported by the Office of the Research and Development of Surigao del Sur State University, Main Campus, Tandag City Surigao del Sur, Philippines.

REFERENCES

[1] Z. Masood, R. Samar, M. A. Z. Raja, "Design of a mathematical model for the Stuxnet virus in a network of critical control infrastructure," *Computer and Security*, vol. 87, p. 101565, 2019, doi: 10.1016/j.cose.2019.07.002.
[2] P. Shahrear, A. K. Chakraborty, A. Islam, U. Habiba, "Analysis of Computer Virus Propagation Based on Compartmental Model," *Applied and Computational Mathematics*, vol. 7, no. 1, pp. 12-21, 2017, doi: 10.11648/j.acm.s.2018070102.12.
[3] A. Raza *et al*, "Mathematical analysis and design of the nonstandard computational method for an epidemic model of computer virus with delay Effect: Application of mathematical biology in computer science," *Results in Physics*, vol. 21, p. 103750, 2021, doi: 10.1016/j.rinp.2020.103750.
[4] G. Maji, S. Mandal, S. Sen, "A systematic survey on influential spreaders identification in complex networks with a focus on K-shell based techniques," *Expert Systems with Applications*, vol. 161, p. 113681, 2020, doi: 10.1016/j.eswa.2020.113681.
[5] R. K. Upadhyay, P. Singh, "Modeling and control of computer virus attack on a targeted network," *Physica A: Statistical Mechanics and its Applications*, vol. 538, p. 122617, 2020, doi: 10.1016/j.physa.2019.122617.
[6] L. X. Yang, X. Yang, "The spread of computer viruses over a reduced scale-free network," *Physica A: Statistical Mechanics and its Applications*, vol. 396, pp. 173-184, 2014, doi: 10.1016/j.physa.2013.11.026.
[7] W. Tang, *et al*, "SLBRS: Network Virus Propagation Model based on Safety Entropy," *Applied Soft Computing*, vol. 97, p. 106784, 2020, doi: 10.1016/j.asoc.2020.106784.
[8] R. Dubey, S. Bharadwaj, M. I. Zafar, V. Bhushan Sharma, S. Biswas, "Collaborative noise mapping using smartphone," *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 43, no. B4, pp. 253-260, 2020, doi: 10.5194/isprs-archives-XLIII-B4-2020-253-2020.
[9] O. Bamaarouf, A. Ould Baba, S. Lamzabi, A. Rachadi, H. Ez-Zahraouy, "Effects of maximum node degree on computer virus spreading in scale-free networks," *International Journal of Modern Physics B*, vol. 31, no. 26, pp. 1-10, 2017, doi: 10.1142/S021797921750182X.
[10] C. Gan, Q. Feng, Q. Zhu, Z. Zhang, Y. Zhang, Y. Xiang, "Analysis of computer virus propagation behaviors over complex networks: a case study of Oregon routing network," *Nonlinear Dynamics*, vol. 100, no. 2, pp. 1725-1740, 2020, doi: 10.1007/s11071-020-05562-1.
[11] A. Bhargava, D. K. Soni, P. Jain, J. Dhar, "Dynamics of attack of malicious codes on the targeted network: Effect of firewall," *International Conference on Recent Trends in Information Technology (ICRTIT)*, 2016, pp. 1-6, doi: 10.1109/ICRTIT.2016.7569534.
[12] L. X. Yang, X. Yang, "The pulse treatment of computer viruses: A modeling study," *Nonlinear Dynamics*, vol. 76, no. 2, pp. 1379-1393, 2014, doi: 10.1007/s11071-013-1216-x.
[13] L. Long, K. Zhong, W. Wang, "Malicious viruses spreading on complex networks with heterogeneous recovery rate," *Physica A: Statistical Mechanics and its Applications*, vol. 509, pp. 746-753, 2018, doi: 10.1016/j.physa.2018.05.149.
[14] B. Zhang, L. Zhang, C. Mu, Q. Zhao, Q. Song, X. Hong, "A most influential node group discovery method for influence maximization in social networks: A trust-based perspective," *Data & Knowledge Engineering*, vol. 121, pp. 71-87, 2019, doi: 10.1016/j.datak.2019.05.001.
[15] S. Kumari, P. Singh, R. K. Upadhyay, "Virus dynamics of a distributed attack on a targeted network: Effect of firewall and optimal control," *Communications in Nonlinear Science and Numerical Simulation*, vol. 73, pp. 74-91, 2019, doi: 10.1016/j.cnsns.2019.02.006.
[16] K. S. Kim, M. M. Ibrahim, I. H. Jung, S. Kim, "Mathematical analysis of the effectiveness of control strategies to prevent the autorun virus transmission propagation," *Applied Mathematics and Computation*, vol. 371, p. 124955, 2020, doi: 10.1016/j.amc.2019.124955.
[17] L. Chen, K. Hattaf, J. Sun, "Optimal control of a delayed SLBS computer virus model," *Physica A: Statistical Mechanics and its Applications*, vol. 427, pp. 244-250, 2015, doi: 10.1016/j.physa.2015.02.048.
[18] M. Zhang, G. Song, L. Chen, "A state feedback impulse model for computer worm control," *Nonlinear Dynamics* vol. 85, no. 3, pp. 1561-1569, 2016, doi: 10.1007/s11071-016-2779-0.
[19] L. X. Yang, X. Yang, "A novel virus-patch dynamic model," *PLoS One*, vol. 10, no. 9, pp. 1-16, 2015, doi: 10.1371/journal.pone.0137858.
[20] J. R. C. Piqueira, V. O. Araujo, "A modified epidemiological model for computer viruses," *Applied Mathematics and Computation*, vol. 213, no. 2, pp. 355-360, 2009, doi: 10.1016/j.amc.2009.03.023.
[21] P. Qin, "Analysis of a model for computer virus transmission," *Mathematical Problems in Engineering*, vol. 2015, ID. 720696, 2015, doi: 10.1155/2015/720696.

[22] X. Zhang, "Modeling the spread of computer viruses under the effects of infected external computers and removable storage media," *International Journal of Security and Its Applications*, vol. 10, no. 3, pp. 419-428, 2016, doi: 10.14257/ijsia.2016.10.3.36.
[23] B. K. Mishra, "Mathematical Model on Attack of Worm and Virus in Computer Network," *International Journal of Future Generation Communication and Networking*, vol. 9, no. 6, pp. 245-254, 2016, doi: 10.14257/ijfgcn.2016.9.6.23.
[24] Y. Yao, Q. Fu, W. Yang, Y. Wang, C. Sheng, "An Epidemic Model of Computer Worms with Time Delay and Variable Infection Rate," *Security Communication Networks*, vol. 2018, ID. 9756982, doi: 10.1155/2018/9756982.
[25] S. Tisue, U. Wilensky, "Netlogo: A simple environment for modeling complexity," *Presented at the International Conference on Complex Systems*, pp. 1-10, 2004, [Online]. Available: http://ccl.sesp.northwestern.edu/papers/netlogo-iccs2004.pdf.
[26] E. Sklar, "Software review: NetLogo, a multi-agent simulation environment," *Artif. Life*, vol. 13, no. 3, pp. 303-311, 2007, doi: 10.1162/artl.2007.13.3.303.
[27] J. Ren, X. Yang, Q. Zhu, L. X. Yang, C. Zhang, "A novel computer virus model and its dynamics," *Nonlinear Analysis: Real World Applications*, vol. 13, no. 1, pp. 376-384, 2012, doi: 10.1016/j.nonrwa.2011.07.048.
[28] J. Amador, "The SEIQS stochastic epidemic model with external source of infection," *Applied Mathematical Modelling*, vol. 40, no. 19-20, pp. 8352-8365, 2016, doi: 10.1016/j.apm.2016.04.023.
[29] N. Scaife, H. Carter, P. Traynor, K. R. B. Butler, "CryptoLock (and Drop It): Stopping Ransomware Attacks on User Data," *IEEE 36th International Conference on Distributed Computing Systems (ICDCS)*, 2016, pp. 303-312, doi: 10.1109/ICDCS.2016.46
[30] Y. Yang, *et al*, "General Theory of security and a study of hacker's behavior in big data era," *Peer-to-Peer Networking and Applications*, vol. 11, no. 2, pp. 210-219, 2018, doi: 10.1007/s12083-016-0517-5.

BIOGRAPHY OF AUTHOR

Catherine R. Alimboyong is an Associate Professor at Surigao Del Sur State University, where she teaches computer science. She obtained an MIT degree in 2018 from Saint Paul Univesrity Philippines, Tuguegaro City and a doctorate of IT in 2019 from Technological Institute of Technology, Quezon City, Philippines. Deep learning, network protection, data mining, and IT software project management are some of her current research interests.
cralimboyong@sdssu.edu.ph