International Journal of Technical Innovation in Modern Engineering & Science (IJTIMES)
Impact Factor: 5.22 (SJIF-2017), e-ISSN: 2455-2585
Volume 4, Issue 6, June-2018

EFFICIENT WAY OF RECOGNIZING AND HANDLING MALWARE PROPAGATION IN LARGE SCALE NETWORKS

MRS. SHWETHA K S
Assistant Professor, Dept. of I.S.E, New Horizon College of Engineering, Bangalore, India.

ABSTRACT: ***There have been significant efforts in the field of cyber security to understand the malware and try to break down the bad effects that are associated with a malware. Quite recently, the extent of destruction through the WannaCry Ransomware was seen throughout the world where mission critical and highly confidential data was held at stake. Understanding the complex code of the malware is quite a tedious a task and cannot be done before hand as each malware varies quite differently, so instead, in this idea what we are proposing is to monitor the propagation of a malware and analyze the weak points in the network structure so that this can be modified and improved to prevent and delay the propagation of a malware so that the recovery systems have enough time to ensure that the infected systems return to their normal state.***

KEYWORDS: Malware propagation, Large-scale networks, Malware, SIR model

I INTRODUCTION

Malware are malicious software programs that when deployed by cyber attackers; compromise the security of the computer systems by exploiting their vulnerabilities. The cyber attackers exhaust all their resources into creating complex malware, which can be used to compromise as many networked computers as possible, often motivated by extraordinary financial or political rewards. The malware propagates through a network and compromises systems by exploiting their vulnerabilities, these computer systems that are compromised are called as Bots. These compromised computer systems (Bots) together form the Botnet. These Botnets act as the foundation on which is used by the cyber attackers to attack and compromise further through the network. This poses as a critical challenge to the cyber attackers. [1] With the growth of Internet, more and more computer systems are being connected over a network. This makes the modern society and the people ever more dependent on the global communication medium. Although this improves connectivity and solves many problems, at the same time, criminals and cyber attackers increasingly use the Internet as means to propagate malware and other malicious services. There has also been an emergence of a large black market where hackers or others with criminal intent can purchase malware or use malicious services for a renting fee. These acts as a business, where hackers tend to improve and increase the complexity of the malware to avoid detection by any anti-virus programs. A more complex malware tend to get a higher price, hence providing a strong incentive to the hackers and criminals to ensure complexity and avoid detection. This leads to multiple forks or new implementations of the same type of malicious software that can propagate out of control. The different types of malware that usually propagate through a network are:

Trojan: includes another hidden program, which performs malicious activity in the background.

Potentially Unwanted Program: is usually downloaded together with a freeware program without the user's consent, e.g. toolbars, search engines and games.

Adware: aims at displaying commercials based on the user's information.

Rootkit: has the capability to obfuscate information like running processes or network connections on an infected system. [2]

We find the malware distribution in terms of networks varies from exponential to power law with a short exponential tail, and to power law distribution at its early, late, and final stage, respectively. These findings are firstly theoretically proved based on the proposed model, and then confirmed by the experiments through the two large-scale real-world data sets.

II RELATED WORK

Malware nowadays is spreading widely through the networks and hence poses a critical threat to network security. However, we have very limited understanding on the malware behaviour and about their propagation through a network. We try to understand how a malware propagates through a network on a global perspective. To do so we try to understand a two layer epidemic model, through which we try to explain the malware propagation in a network. Based on the proposed model, the analysis indicates that the distribution of a given malware follows exponential distribution, power law distribution with a short exponential tail, and power law distribution at its early, late and final stages, respectively. The detection of malware propagation in a network in the early stages is extremely difficult. By analyzing and understanding the malware propagation behavior, we will try to contain and quarantine the malware during the late and final stages. We intend to build a simulation model that tries to simulate the propagation of heterogeneous malware in a complex network to aid in the understanding of the spread.

III LITERATURE SURVEY

Malwares for short have become a major security threat. While originating in criminal behavior, their impact are also influenced by the decisions of legitimate end users. Getting agents in the Internet, and in networks in general, to invest in and deploy security features and protocols is a challenge, in particular because of economic reasons arising from the presence of network externalities. Our goal in this paper is to model and quantify the impact of such externalities on the investment in security features in a network. We study a network of interconnected agents, which are subject to epidemic risks such as those caused by propagating viruses and worms. Each agent can decide whether or not to invest some amount to self-protect and deploy security solutions which decreases the probability of contagion. Borrowing ideas from random graphs theory, we solve explicitly this 'micro'- model and compute the fulfilled expectations equilibria. We are able to compute the network externalities as a function of the parameters of the epidemic. [5]

Malicious software has become a major threat to modern society, not only due to the increased complexity of the malware itself but also due to the exponential increase of new malware each day. This study tackles the problem of analyzing and classifying a high amount of malware in a scalable and automatized manner. We have developed a distributed malware testing environment that was used to test an extensive number of malware samples and trace their behavioral data. The extracted data was used for the development of a novel type classification approach based on supervised machine learning. [6]

The paper presents results of a study of malware spreading in heterogeneous networks using epidemiological modeling framework. The model is one of the first to incorporate heterogeneity among the three components of the network: software, hardware and network type. The unified approach taken in this study aggregates and extends models of malware spreading that either do not account for network heterogeneity or allow for heterogeneity within one component, e.g. software. [3]

Malicious software has become a major threat to modern society, not only due to the increased complexity of the malware itself but also due to the exponential increase of new malware each day. This study tackles the problem of analyzing and classifying a high amount of malware in a scalable and automatized manner. We have developed a distributed malware testing environment that was used to test an extensive number of malware samples and trace their behavioral data. The extracted data was used for the development of a novel type classification approach based on supervised machine learning. [7]

IV EPIDEMIC MODEL

SIR Model was introduced by Rozenberg and this model mainly focuses on the worm propagation on the email and social networks.There are two main strategies in email worms propagation, 1)Reinfection 2) Non-reinfection.Reinfection strategy defines whenever any healthy or infected recipients open the malicious attached file the modern email malware sends its copy to the recipients contact. But in non-reinfection strategy the infected user sends the worm to his neighbours only once. This SIR model describes the non-reinfection strategy of the email worm. This epidemic spreading model categorize the population into three states: Susceptible (S) - individuals that are vulnerable and can possibly be infected. Infected (I) - individuals that have been infected and infect other individuals. Removed (R) - infected individuals that do not infect other individuals. For homogeneous networks, dynamics of propagation of email worms can be approximated by SIR model for homogeneous network.

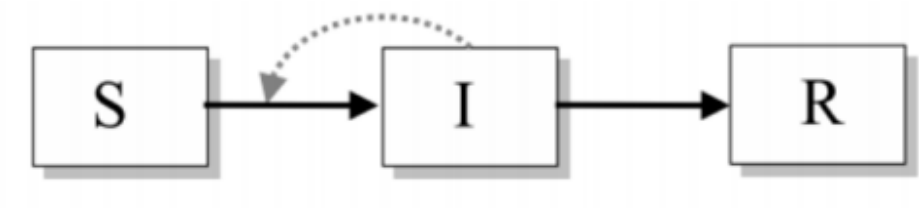

Figure 1: A flow diagram representing the propagation between the different stages in the SIR epidemic model

V WORKING

The program follows the flow as shown in Fig.3. We initially create a two-layer network structure that is similar to the structure shown in Fig.2. The top internet layer by default is created as a ring topology, and the user is then allowed to choose a topology for the underlying network level. We use a mathematical model that considers a lot of network variables such as connectivity, resistance, recovery rate, etc. The mathematical model considers the value of these variables and calculates a value that corresponds to the probability of the host to be infected. We try to analyze how the malware propagation is affected based on the topology as well. For example, a more robust and strongly connected network will aide in the faster propagation of malware as compared to a less robust and weakly connected network. We in this idea are not trying to detect and handle a malware, as that falls in to the realm of cyber security, what we are trying to do instead is to understand how the network structure and different factors of the network affect the propagation of the malware in the network. By understanding the impact of the network structure and other factors, we can modify the network initially at the time of installation to help delay and handle the propagation of malware.

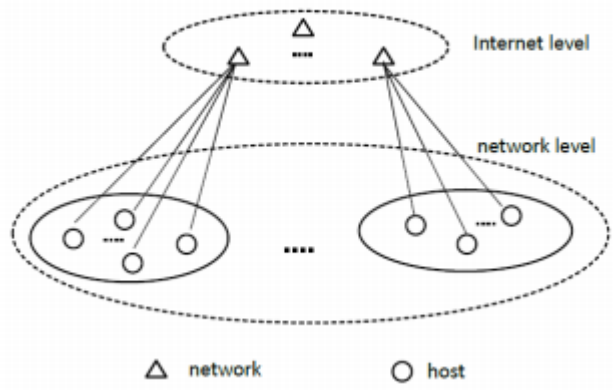

Figure 2: The two layer network structure used to create the large-scale network for studying malware propagation.

VI PROPOSED METHOD

This proposed idea is being implemented using Spring Boot framework that is built as bootstrap framework on top of Java that simplifies most of the mundane tasks using annotations. The program is run as a standard java application. The user is allowed to choose the size and underlying network topology initially when the program is executed. The user enters the number of internet level and network level nodes respectively and chooses the underlying network topology. Once the values have been accepted from the user, the program introduces the malware in one the leaf nodes of a host in the network level. Once the malware has been introduced, we start the timer, which is used to calculate the total time taken for the malware to propagate. The host on which the malware was initially introduced now starts attacking all of its direct children and its parent node based on the network topology. In the case of ring topology, it directly attacks the immediate neighbours in the network.

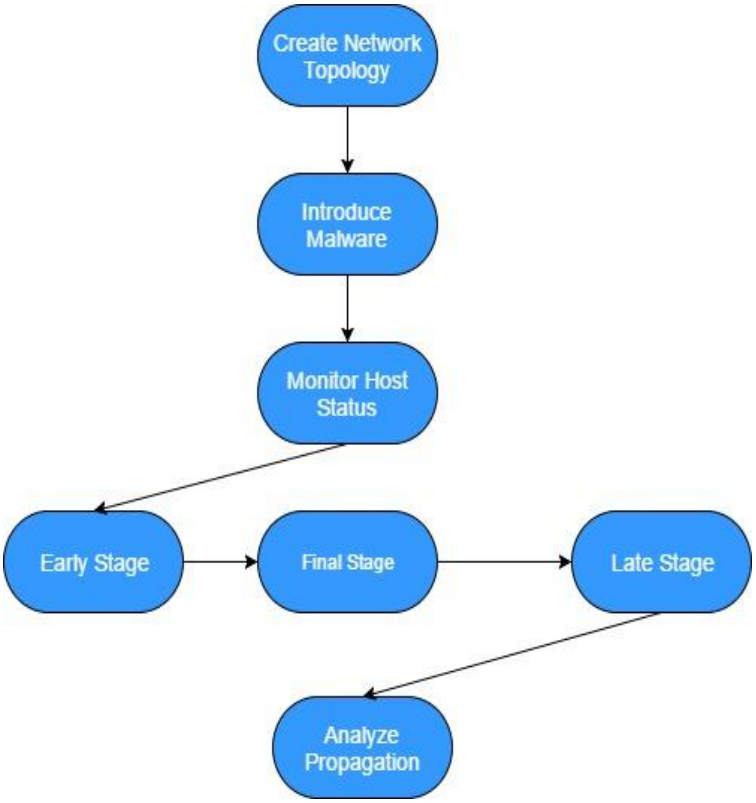

Figure 3: The flow of how the idea is executed and malware propagation is analyzed.

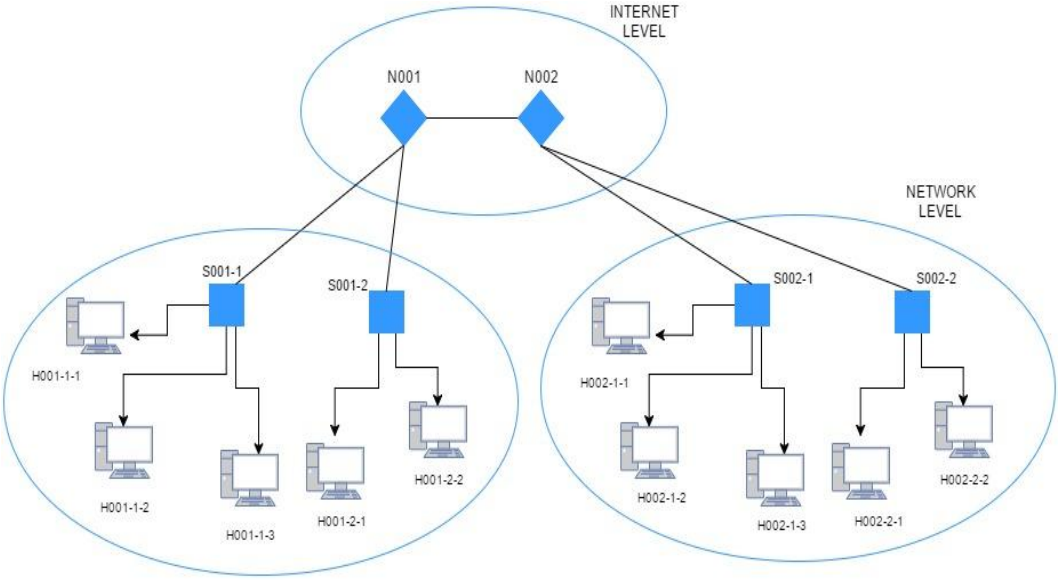

Figure 4: An two layer network structure with two internet level nodes and seven network level nodes with an underlying network with star topology.

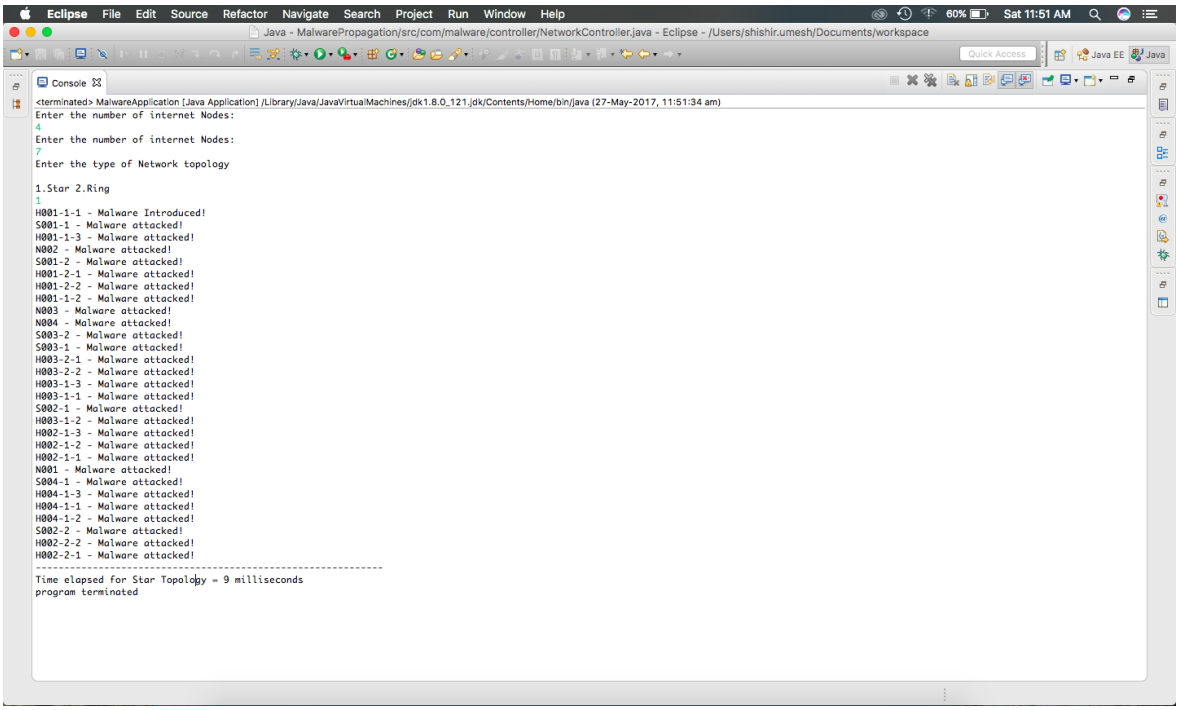

Figure 5: Output of a propagation with four internet level nodes and seven network level nodes for a underlying network level star topology.

VII SIMULATION RESULTS

This simulation is repeated several times to for the same and different topologies for the same number of nodes in the large-scale network and time values are noted down. Figure 5 shows the execution of the program where there are four internet level nodes and seven network level nodes with an underlying network with star topology. The

Start Active time	Alarm	Source	Source - Group	Target	Details
50 milliseconds	Malware Worm Propagation	-	S001-1	H001-1-1	Malware Worm Introduced
95 milliseconds	Malware Worm Propagation	H001-1-1	N001	S001-1	Malware Worm propagated using network LAN
130 milliseconds	Malware Worm Propagation	S001-1	S001-1	H001-1-3	Malware Worm propagates using network LAN

Table 1: This table is a small excerpt that shows the propagation of malware through the network and all the necessary details for monitoring and understanding

naming convention used within the program and in the output in Figure 5 are similar to the topology and naming convention used in Figure 4. The program can be run for ring topology as well.

The table 1 is a representation of the malware propagation, which is shown as a screenshot shown in figure 5. This simulation was repeatedly for the input values of fifteen network level nodes with each internet level node having forty network level hosts under it, connected to the network level in a star topology. The average time taken in the default network structure that was created was 372 milliseconds. The same network structure with an underlying ring topology resulted in an average simulation time of 453 milliseconds.

The aim behind the idea is to ensure that we are able to delay the time it takes to reach the final stage hence ensuring that the hosts have enough time to recover from their infected stage. By increasing the resistance of star hub nodes and the internet level nodes, now running the simulation repeatedly resulted in an average time of 550 milliseconds for the malware propagation. This shows how simple modifications to the network structure, network attributes and host attributes can help in delaying the time taken for malware propagation. Although the time cannot be compared to the real life scenario, it can be thought to be directly proportional, i.e. increasing the delay from "x" milliseconds to "x + constant" milliseconds in the program can be thought of delaying the propagation in real time scenario.

VIII CONCLUSION

In this paper we conclude that understanding the propagation of a malware is critical to prevent and handle the spread of malware through a network. Therefore, by repeated experiments and simulations it is clear that the structure of the network and different network attributes can be modified to ensure that we can delay the propagation of the malware. We can further work towards developing a mathematical model that is more accurate and which through the consideration of recovery rate will be able to give better and deeper insights into the propagation of a malware in the large scale network.

REFERENCES

[1] Shui YU, Guofei Gu, Ahmed Barnavi, Song Guo, Ivan Stojmenovic, "Malware Propagation in Large-Scale Networks", IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING , pp. 170 – 179, 2014.

[2] Kayla.M.Straub, Avik Sengupta, Joseph.M.Ernst, Robert W. McGwier, Merrick Watchorn, Richard Tilley, Randolph Marchany, "Malware Propagation in Fully Connected Networks : A Net flow-Based Analysis", Military Communications Conference, MILCOM 2016 - 2016 IEEE.

[3] Alexander Alexeev, Diana S. Henshel, Mariana Cains, Quan Sun, "On the Malware Propagation in Heterogeneous Networks", 2016 IEEE 12th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob).

[4] Bo Liu, Wanlei Zhou, Longxiang Gao, HaiBo Zhou, Tom H. Luan, Sheng Wen, "Malware Propagation in Wireless Ad Hoc Networks", IEEE Transactions on Dependable and Secure Computing, pp. 1-1, 2016.

[5] Marc Lelarge, "Economics of Malware: Epidemic Risks Model, Network Externalities and Incentives", 2009 47th Annual Allerton Conference on Communication, Control, and Computing (Allerton), 2009.

[6] Radu S. Pirscoveanu, Steven S. Hansen, Thor M. T. Larsen, Matija Stevanovic, Jens Myrup Pedersen, Alexandre Czech, "Analysis of Malware Behaviour: Type Classification using Machine Learning", 2015 International Conference on Cyber Situational Awareness, Data Analytics and Assessment (CyberSA), 2015.

[7] Marcus Martens, Hadi Asghari, Michel van Eeten, Piet Van Mieghem, "A Time-dependent SIS-model for Long-term Computer Worm Evolution", 2016 IEEE Conference on Communications and Network Security (CNS), 2016.

[8] Dhende Kapil N., Prof. Bere S. S, "A Review on: Malware Propagation in Large Scale Networks", International Journal of Innovative Research in Computer and Communication Engineering, Vol. 3, Issue 11, November 2015.

[9] Wenzhi Chen, "A Mathematical Model of Ebola Virus Based on SIR Model", 2015 International Conference on Industrial Informatics - Computing Technology, Intelligent Technology, Industrial Information Integration, 2015.

[10] P. V. Mieghem, J. Omic, and R. Kooij, "Virus spread in networks," IEEE/ACM Trans. Netw., vol. 17, no. 1, pp. 1–14, Feb. 2009.

[11] https://en.wikipedia.org/wiki/Epidemic_model

[12] https://en.wikipedia.org/wiki/Probability_distribution

[13] https://en.wikipedia.org/wiki/Malware