ISSN(Online) : 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

An ISO 3297: 2007 Certified Organization Vol.5, Special Issue 5, June 2017

8th One Day National Conference on Innovation and Research in Information Technology (IRIT- 2017)

Organized by

Departments of ISE, CSE & MCA, New Horizon College of Engineering, Bengaluru, Karnataka 560103, India

Recognizing and Handling the Malware Propagation in Large Scale Networks

Rashmi Priya[1], Shishir Umesh[1], Shwetha K S[2], Dr. Jitendranath Mungara[3]

Student, Dept. of I.S.E, New Horizon College of Engineering, Bangalore, India.[1]

Assistant Professor, Dept. of I.S.E, New Horizon College of Engineering, Bangalore, India.[2]

Head of Department, Dept. of I.S.E, New Horizon College of Engineering, Bangalore, India.[3]

ABSTRACT: 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 behavior 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.

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. [2] 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.

MALWARE EMERGENCE AND PROPAGATION

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. This 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:

ISSN(Online) : 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

An ISO 3297: 2007 Certified Organization Vol.5, Special Issue 5, June 2017

8th One Day National Conference on Innovation and Research in Information Technology (IRIT- 2017)

Organized by

Departments of ISE, CSE & MCA, New Horizon College of Engineering, Bengaluru, Karnataka 560103, India

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. [6]

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. 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]

ISSN(Online) : 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

An ISO 3297: 2007 Certified Organization Vol.5, Special Issue 5, June 2017

8th One Day National Conference on Innovation and Research in Information Technology (IRIT- 2017)

Organized by

Departments of ISE, CSE & MCA, New Horizon College of Engineering, Bengaluru, Karnataka 560103, India

III. PROPOSED MODEL

EPIDEMIC MODELS

We use an epidemic model, as a simplified mean to describe the propagation and transmission of communicable diseases through individuals on a global perspective. One common type of epidemic model is the SIR (Susceptible-Infected-Recovered) Model. This model categorizes the hosts within a population as **Susceptible** (if previously unexposed to the disease), **Infected** (if currently colonized by the pathogen) and **Recovered** (if they had successfully cleared the infection). An individual in the population progresses from S to I if it involves disease transmission, which is determined by three distinct factors; the prevalence of the infection or disease, the network or contact structure of the underlying population and the probability of transmission given there is contact among the underlying population. For a pathogen of the disease to be directly transmitted, there has to be contact between the susceptible and infected individuals and the probability of this happening is determined by respective levels of S and I, as well as the contact structure of inherent host population. Finally, we need to take into account the likelihood that a contact between a susceptible and an infectious person results in transmission. Figure 1 shows the movement between the S and I classes and the I and R classes as black arrows. [9]

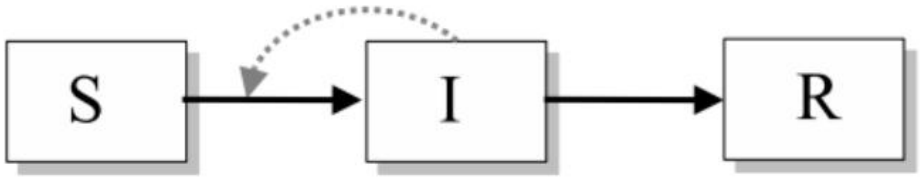

Fig. 1. A flow diagram representing the movement of a body between the different stages in the SIR epidemic model

DISTRIBUTION MODELS (PROPOSED)

In probability theory and statistics, the exponential distribution (also known as negative exponential distribution) is the probability distribution that describes the time between events in a Poisson process, i.e. a process in which events occur continuously and independently at a constant average rate. It is a particular case of the gamma distribution. It is the continuous analogue of the geometric distribution, and it has the key property of being memoryless. In addition to being used for the analysis of Poisson processes, it is found in various other contexts. In statistics, a power law is a functional relationship between two quantities, where a relative change in one quantity results in a proportional relative change in the other quantity, independent of the initial size of those quantities: one quantity varies as a power of another. For instance, considering the area of a square in terms of the length of its side, if the length is doubled, the area is multiplied by a factor of four.

PROPOSED EPIDEMIC MODEL

Figure 2 shows the proposed two layer epidemic model that will be used to study the propagation of malware across the networks.The proposed model has two layers: An internet layer and a network layer. The internet layer is the top layer the connects the multiple underlying network layers. The cyber attackers or criminals use the internet layer to introduce the malware and other malicious software into the network. The malware then propagates through the network compromising computer system and creating its Botnet.

ISSN(Online) : 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

An ISO 3297: 2007 Certified Organization Vol.5, Special Issue 5, June 2017

8th One Day National Conference on Innovation and Research in Information Technology (IRIT- 2017)

Organized by

Departments of ISE, CSE & MCA, New Horizon College of Engineering, Bengaluru, Karnataka 560103, India

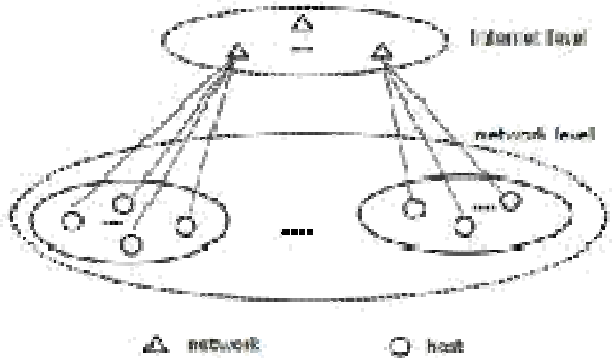

Fig.2 This shows the proposed two layer epidemic model which will be used to study the malware propagation in a network

IV. CONCLUSION

In this paper, we try to thoroughly explore the problem of malware propagation and distribution at large-scale networks. Malware poses a critical challenge and threat to cyber security and requires immediate attention and a probable solution to the problem of Malware propagation. The epidemic model consists of two layers: the internet/upper layer focuses on networks of a large-scale connectivity or the internet whereas the network/lower layer focuses on the connectivity between physicals hosts of a given network. This not just helps in the separation of concerns, but improves the understanding of the way Malware propagates on a global perspective. Moreover, the proposed two layer model offers us the distribution of malware in terms of the low layer networks. An analysis on the malware propagation on the epidemic model results in the following conclusion: Complex networks have demonstrated that the number of hosts of networks follows the power law. People found that the size distribution usually follows the power law, such as population in cities in a country or personal income in a nation. In terms of the Internet, researchers have also discovered many power law phenomenon's, such as the size distribution of web files. Recent progresses reported in further demonstrated that the size of networks follows the power law. The power law has two expression forms: the Pareto distribution and the Zipf distribution. For the same objects of the power law, we can use any one of them to represent it. However, the Zipf distributions are tidier than the expression of the Pareto distributions. In this paper, we will use Zipf distributions to represent the power law.

REFERENCES

[1] Malware Propagation in Large-Scale Networks by Shui YU, Guofei Gu, Ahmed Barnavi, Song Guo, Ivan Stojmenovic
[2] Malware Propagation in Fully Connected Networks : A Net flow-Based Analysis
[3] On the Malware Propagation in Heterogeneous Networks by Alexander Alexeev, Diana S.Henshel, Mariana Cain and Quan Sun
[4] Malware Propagation in Wireless Ad Hoc Networks by Bo Liu, Tom H.Luans
[5] Economics of Malware: Epidemic Risks Model, Network Externalities and Incentives by Marc Lelarge
[6] Analysis of Malware Behaviour: Type Classification using Machine Learning
[7] A Time-dependent SIS-model for Long-term Computer Worm Evolution by Marcus Martens, Hadi Asghari, Michel Van Eeten and Piet Van Mieghem
[8] https://en.wikipedia.org/wiki/Epidemic_model
[9] https://en.wikipedia.org/wiki/Epidemic_model
[10] https://en.wikipedia.org/wiki/Malware