Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X
IMPACT FACTOR: 6.017

IJCSMC, Vol. 6, Issue. 4, April 2017, pg.173 – 179

LITERATURE REVIEW ON MALWARE PROPAGATION IN LARGE SCALE NETWORKS

Kavitha.U, S.Shanmugapriya

Student, Senior Assistant Professor

M.Tech, Dept. of CSE, New Horizon College of Engineering

I. INTRODUCTION

Malware is the term assigned to the malicious software programs that are deployed by cyber attackers to compromise the computer systems by exploiting their security vulnerabilities. Being motivated by the extraordinary financial or political rewards, malware owners exhaust their energy to compromise as many networked computers as they can with an intention to achieve their malicious goals. A compromised computer is called a bot, and all bots compromised by a malware result in formation of a botnet. Botnets have emerged as the attack engine of cyber attackers, and they pose critical challenges to cyber defenders. It is important for defenders to understand malware behavior, such as propagation or membership recruitment patterns, the size of botnets, and distribution of bots so that they can fight against cyber criminals[1].

Till today, we do not have a clear understanding about the size and distribution of malware or botnets. There are various methods that are employed by the researchers to measure the size of botnets, such as botnet infiltration DNS redirection external information[1]. These efforts indicate that the size of botnets varies from millions to a few thousand. There are no principles that are dominant to explain these variations. As a result, researchers desperately desire effective models and explanations for the chaos. Dagon et al. revealed that time zone has an obvious impact on the number of available bots[2]. Mieghem et al. indicated that network topology has an important impact on malware spreading through their rigorous mathematical analysis.

II. LITERATURE SURVEY

The basic flow of malware propagation is as follows. A malware programmer writes a program, which will be called as a bot or agent, and then he installs these bots at the compromised computers on the Internet using various network virus-like techniques. All of his bots then result in formation of a botnet, which are controlled by its owners to commit illegal tasks, such as launch of DDoS attacks, sending spam emails, performation of phishing activities, and collection of sensitive information[3]. There exists a command and control (C&C) server(s) that communicates with the bots and then collect data from bots. The botmaster changes the url of his C&C frequently in order to disguise himself from legal forces, e.g., weekly[4]. An excellent explanation about this can be found in with the significant growing of smartphones, we have witnessed an increasing number of mobile malware. Malware writers have developed several mobile malware in last few years[5].

Some of the previous works related to analysis of malware:

1) Information-theoretic view of network aware malware attacks

Smartphones are pervasively used in society, and have been both the target and victim of malware writers. In this approach, we survey the current smartphone malware status and their propagation models, being motivated by the significant threat that presents to legitimate users, [6]. The content of this paper is presented in two parts. In the beginning of first part, we perform the review of short history of mobile malware evolution since 2004, and then they list the classes of mobile malware and their infection vectors[6]. At the end of the first part, we list the damage that is possibly caused by smartphone malware. In the whole of second part, we focus on smartphone malware propagation modeling[6]. We recall generic epidemic models as a foundation for further exploration in order to understand the propagation behavior of smartphone malware,. We then extensively survey the smartphone malware propagation models[6].

Disadvantage:

- It only discusses the behavior of malwares.

2) Modeling and automated containment of worms

Self-propagating codes which are also usually termed as worms, some of which like Code Red, Nimda, and Slammer, have drawn significant attention due to their enormously adverse impact on the Internet. Therefore, there is a huge interest in the research community for modeling the spread of worms and in providing adequate defense mechanisms against them. In this paper, we present a (stochastic) branching process model for characterizing the propagation of Internet worms[7]. The model is developed for uniform scanning of worms and then it is extended to preference scanning worms. This model leads to the development of an containment strategy that enables the prevention of the spread of a worm beyond its early stage[7].

Also Specifically for the purpose of uniform scanning of worms, we are able to:

1) Provide a condition that is precise and it determines whether the spread of the worm will eventually stop

2) Obtain the distribution of the total number of hosts that the worm infects.

We then extend our results to contain preference scanning worms[7]. Our strategy is based on limiting the number of scans to dark-address space. The limiting value is determined by our analysis. Our automatic worm containment schemes effectively contain both uniform scanning worms and local preference scanning worms, and it is validated through simulations and real trace data to be nonintrusive[7].

Disadvantage:

- It is not possible to prevent undesired messages. No matter user who propose them.

3) An epidemic theoretic framework for vulnerability analysis of broadcast protocols in wireless sensor networks

While multi-hop broadcast protocols, namely Trickle, Deluge and MNP, have gained tremendous popularity as a means for fast and convenient propagation of data/code in large scale wireless sensor networks, however, unfortunately, they can serve as potential platforms for spreading of viruses under the condition if the security is breached. To understand the vulnerability of such protocols and design defense mechanisms against piggy-backed virus attacks, it is critical to investigate the propagation process of these protocols in terms of their speed and reachability.

In this paper, we propose a general framework based on the principles of epidemic theory, for vulnerability analysis of current broadcast protocols in wireless sensor networks[8]. In particular, we develop a common mathematical model for the propagation that incorporates important parameters derived from the communication patterns of the protocol under test. Based on this model, analysis of the propagation rate and the extent of spread of a malware over typical broadcast protocols is being done as proposed in the literature[8]. The overall result is an approximate but convenient tool to characterize a broadcast protocol in terms of its vulnerability to malware propagation.

Disadvantage:

- It uses the access control techniques to block Malware.

4) A large-scale empirical study of conficker

Conficker is the most recent widespread, well-known worm/bot. According to several reports, it has infected about 7 million to 15 million hosts and the victims are still increasing even now. In this paper, we analyze Conficker infections at a large scale, about 25 million victims, and study various interesting aspects about this state-of-the-art malware. By analyzing Conficker, we intend to understand current and new trends in malware propagation, which could be very helpful in predicting future malware trends and providing insights for future malware defense.

We observe that Conficker has some very different victim distribution patterns compared to many previous generation worms/botnets, suggesting that new malware spreading models and defense strategies are likely needed. We measure the potential power of Conficker to estimate its effects on the networks/hosts when it performs malicious operations[9]. Furthermore, we intend to determine how well a reputation-based blacklisting approach can perform when faced with new malware threats such as Conficker.

We cross-check several DNS blacklists and IP/AS reputation data from Dshield and FIRE and our evaluation shows that unlike a previous study which shows that a blacklist-based approach can detect most bots, these reputation-based approaches did relatively poorly for Conficker[9]. This raises a question of how we can improve and complement existing reputation-based techniques to prepare for future malware defense? Based on this, we look into some insights for defenders. We show that neighborhood watch is a surprisingly effective approach in the case of Conficker.

Disadvantage:

- By Providing this service, it involves the matter of using previously defined web content mining techniques for a different application, along with that, it also requires to design ad-hoc classification strategies.

III. PROPOSED SYSTEM

In this paper, we use two large scale malware data sets for our experiments. Conficker is a well-known and one of the most recently widespread malware. Shin et al. collected a data set about 25 million Conficker victims from all over the world at different levels. At the same time, malware targeting on Android based mobile systems are developing quickly in recent years. Zhou and Jiang collected a large data set of Android based malware.

In this paper, we conduct the study of the distribution of malware in terms of networks (e.g., autonomous systems (AS), ISP domains, abstract networks of smartphones who share the same vulnerabilities) at large scales[3]. In this kind of setting, we have a sufficient volume of data at a large enough scale to meet the requirements of the SI model. Different from the traditional epidemic models, we break our model into two layers[3]. First of all, for a given time since the breakout of a malware, we calculate how many networks have been compromised based on the SI model. Second, for a compromised network, we calculate how many hosts have been compromised since the time that the network was compromised. With this two layer model in place, we can determine the total number of compromised hosts and their distribution in terms of networks. Through our rigorous analysis, we find that the distribution of a given malware follows an exponential distribution at its early stage, and obeys a power law distribution with a short exponential tail at its late stage, and finally converges to a power law distribution[3].

The proposed two layer epidemic model and the findings are the first work in the field.
The summary of our contributions are as follows[3]:

- A two layer malware propagation model is proposed to describe the development of a given malware at the Internet level. The proposed model represents malware propagation better in large-scale networks when compared with the existing single layer epidemic models,.
- 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[3]. These findings are first theoretically proved based on the proposed model, and then confirmed by the experiments through the two large-scale real-world data sets.

IV. FUTURE WORK

In regards to future work, we will first further investigate the dynamics of the late stage. More details of the findings are expected to be further studied, such as the length of the exponential tail of a power law distribution at the late stage. Second, defenders may care more about their own network, e.g., the distribution of a given malware at their ISP domains, where the conditions for the two layer model may not hold. We need to seek appropriate models to address this problem. Finally, we are interested in studying the distribution of multiple malware on large-scale networks as we only focus on one malware in this paper. We believe it is not a simple linear relationship in the multiple malware case compared to the single malware one.

REFERENCES

[1] http://omnetsimulation.com/malware-propagation-in-large-scale-networksmalware-propagation-in-large-scale-networksomnet-simulation/

[2] Fool Me If You Can: Mimicking Attacks and Anti-Attacks in Cyberspace, Article · Jan 2015 · IEEE Transactions on Computers

[3] S. Yu, G. Gu, A. Barnawi, S. Guo and I. Stojmenovic, "Malware Propagation in Large-Scale Networks," in *IEEE Transactions on Knowledge and Data Engineering*, vol. 27, no. 1, pp. 170-179, Jan. 1 2015.

[4] IJISET - International Journal of Innovative Science, Engineering & Technology, Vol. 2 Issue 9, September 2015. www.ijiset.com ISSN 2348 – 7968 An Efficient Detection Technique: Malware Spreading in Peer-to-Peer Networks

[5] The Future of Mobile Malware By: Laura O'Brien

[6] S. Peng, S. Yu and A. Yang, "Smartphone Malware and Its Propagation Modeling: A Survey," in *IEEE Communications Surveys & Tutorials*, vol. 16, no. 2, pp. 925-941, Second Quarter 2014.

[7] S. H. Sellke, N.B. Shroff, and S. Bagchi, "Modeling and automated containment of worms," IEEE Trans. Dependable Secure Comput., vol. 5, no. 2, pp. 71–86, Apr.–Jun. 2008.

[8] P. De, Y. Liu, and S. K. Das, "An epidemic theoretic framework for vulnerability analysis of broadcast protocols in wireless sensor networks," IEEE Trans. Mobile Comput., vol. 8, no. 3, pp. 413–425, Mar. 2009

[9] S. Shin, G. Gu, N. Reddy and C. P. Lee, "A Large-Scale Empirical Study of Conficker," in *IEEE Transactions on Information Forensics and Security*, vol. 7, no. 2, pp. 676-690, April 2012.