ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

Detection of Malware Attacks in Large-Scale Networks

T.Ram Babu
M.Tech Student,
Department of CSE,
Sree Rama institute of Technology and Science,
Kuppenakuntla, Penuballi, Khammam,TS India.

N.Naveen
Assistant Professor,
Department of CSE,
Sree Rama institute of Technology and Science,
Kuppenakuntla, Penuballi, Khammam,TS India.

ABSTRACT:

Malware is pervasive in networks, and poses a critical threat to network security. However, we have very limited understanding of malware behavior in networks to date. In this paper, we investigate how malware propagates in networks from a global perspective. We formulate the problem, and establish a rigorous two layer epidemic model for malware propagation from network to network. Based on the proposed model, our 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. Extensive experiments have been performed through two real-world global scale malware data sets, and the results confirm our theoretical findings.

Index Terms:

Malware, propagation, modeling, power law.

INTRODUCTION:

MALWARE are malicious software programs deployed by cyber attackers to compromise computer systems by exploiting their security vulnerabilities. Motivated by extraordinary financial or political rewards, malware owners are exhausting their energy to compromise as many networked computers as they can in order to achieve their malicious goals. A compromised computer is called a bot, and all bots compromised by a malware form a botnet. Botnets have become the attack engine of cyber attackers, and they pose critical challenges to cyber defenders.

In order to fight against cyber criminals, it is important for defenders to understand malware behavior, such as propagation or membership recruitment patterns, the size of botnets, and distribution of bots.

Existing System:

To date, we do not have a solid understanding about the size and distribution of malware or botnets. Researchers have employed various methods to measure the size of botnets, such as botnet infiltration [1], DNS redirection [3], external information [2]. These efforts indicate that the size of botnets varies from millions to a few thousand. There are no dominant principles to explain these variations. As a result, researchers desperately desire effective models and explanations for the chaos. Dagon et al. [3] revealed that time zone has an obvious impact on the number of available bots. Mieghem et al. [4] indicated that network topology has an important impact on malware spreading through their rigorous mathematical analysis. Recently, the emergence of mobile malware, such as Cabir [5], Ikee [6], and Brador [7], further increases the difficulty level of our understanding on how they propagate. More details about mobile malware can be found at a recent survey paper [8]. To the best of our knowledge, the best finding about malware distribution in large-scale networks comes from Chen and Ji [9]: the distribution is non-uniform. All this indicates that the research in this field is in its early stage.

Proposed System:

In this paper, we study the distribution of malware in terms of networks (e.g., autonomous systems (AS), ISP domains,

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

abstract networks of smartphones who share the same vulnerabilities) at large scales. 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. First of all, for a given time since the break-out 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 theirdistribution 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 tailat its late stage, and finally converges to a power law distribution. We examine our theoretical findings through two large-scale real-world data sets: the Android based malware [19] and the Conficker [20]. The experimental resultsstrongly support our theoretical claims. To the best of our knowledge, the proposed two layer epidemic model and the findings are the first work in the field.

RELATED WORK:

The basic story of malware is as follows. A malware programmer writes a program, called bot or agent, and then installs the bots at compromised computers on the Internet using various network virus-like techniques. All of his bots form a botnet, which is controlled by its owners to commit illegal tasks, such as launching DDoS attacks, sending spam emails, performing phishing activities, and collecting sensitive information. There is a command and control (C&C) server(s) to communicate with the bots and collect data from bots. In order to disguise himself from legal forces, the botmaster changes the url of his C&C frequently, e.g., weekly. An excellent explanation about this can be found in [1]. With the significant growing of smartphones, we have witnessed an increasing number of mobile malware. Malware writers have develop many mobile malware in recent years. Cabir [5] was developed in 2004, and was the first malware targeting on the Symbian operating system for mobile devices. Moreover, it was also the first malware propagating via Bluetooth. Ikee [6] was the first mobile malware against Apple iPhones, while Brador [7] was developed against Windows CE operating systems.

The attack victors for mobile malware are diverse, such as SMS, MMS, Bluetooth, WiFi, and Web browsing. Peng et al. [8] presented the short history of mobile malware since 2004, and surveyed their propagation models.A direct method to count the number of bots is to use botnet infiltration to count the bot IDs or IP addresses. Stone- Gross et al. [1] registered the URL of the Torpig botnet before the botmaster, and therefore were able to hijack the C&C server for ten days, and collect about 70G data from the bots of the Torpig botnet. They reported that the footprint of the Torpig botnet was 182,800, and the median and average size of the Torpig's live population was 49,272 and 48,532, respectively. They found 49,294 new infections during the ten days takeover. Their research also indicated that the live population fluctuates periodically as users switch between being online and offline. This issue was also tacked by Dagon et al. in [3].

Another method is to use DNS redirection. Dagon et al. [3] analyzed captured bots by honypot, and then identified the C&C server using source code reverse engineering tools. They then manipulated the DNS entry which is related to a botnet's IRC server, and redirected the DNS requests to a local sinkhole. They therefore could count the number of bots in the botnet. As discussed previously, their method counts the footprint of the botnet, whichwas 350,000 in their report. 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. [20] 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 [19] collected a large data set of Android based malware. In [2], Rajab et al.pointed out that it is inaccurate to count the unique IP addresses of bots because DHCP and NAT techniques are employed extensively on the Internet ([1] confirms this by their observation that 78.9 percent of the infected machines were behind a NAT, VPN, proxy, or firewall). They therefore proposed to examine the hits of DNS caches to find the lower bound of the size of a given botnet.

PRELIMINARIES:

The boundary nodes identified so far are discrete. They largely depict the network boundaries. However, many applications require not only discrete boundarynodes, but also closed boundary surfaces.

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

Moreover, it is highly desirable that such surfaces are locally planarized 2-manifold in order to apply available 2-D graphic tools on 3-D surfaces. In this research, we implement an algorithm that constructs locally planarized triangular meshes on the identified 3-D boundaries. We adopt the method proposed in [30] that can produce a 2-D planar subgraph (which, however, is not a triangular mesh) and extend it to 3-D surfaces to achieve complete triangulation without degenerated edges. The algorithm is localized and based on connectivity only.

TABLE 1
Notations of Symbols in This Paper

Notation	Description
$I(t)$	Number of infected hosts at time t
$R(t)$	Number of recovered hosts at time t
N	The total number of vulnerable hosts
$\beta(t)$	The infection rate at time t
$S(L_i,t)$	Number of infected hosts of network L_i at time t
$L^j_{k_i}$	The j^{th} network compromised at round i

PRELIMINARIES:

Preliminaries of epidemic modelling and complex networks are presented in this section as this work is mainly based on the two fields. For the sake of convenience, we summarize the symbols that we use in this paper in Table 1.

Deterministic Epidemic Models:

After nearly 100 years development, the epidemic models [17] have proved effective and appropriate for a system that possesses a large number of vulnerable hosts. In other words, they are suitable at a macro level. Zou et al. [15] demonstrated that they were suitable for the studies of Internet based virus propagation at the early stage.We note that there are many factors that impact the malware propagation or botnet membership recruitment, such as network topology, recruitment frequency, and connection status of vulnerable hosts. All these factors contribute to the speed of malware propagation. Fortunately, we can include all these factors into one parameter as infection rate b in epidemic theory. Therefore, in our study, let N be the total number of vulnerable hosts of a large-scale network (e.g., the Internet) for a given malware.

Complex Networks:

Research on 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 [24]. In terms of the Internet, researchers have also discovered many power law phenomenon, such as the size distribution of web files [25]. Recent progresses reported in [26] further demonstrated that the size of networks follows the power law.

PROBLEM DESCRIPTION:

In this section, we describe the malware propagation problem in general. As shown in Fig. 2, we study the malware propagation issue at two levels, the Internet level and the network level. We note that at the network level, a network could be defined in many different ways, it could be an ISP domain, a country network, the group of a specific mobile devices, and so on. At the Internet level, we treat every network of the network level as one element.

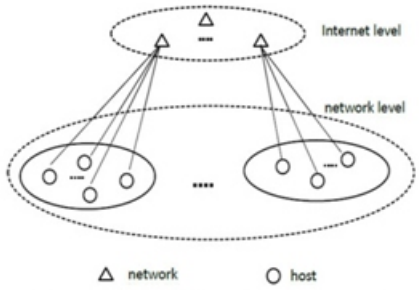

Fig. 2. The system architecture of the studied malware propagation.

PERFORMANCE EVALUATION:

In this section, we examine our theoretical analysis through two well-known large-scale malware: Android malware and Conficker. Android malware is a recent fast developing and dominant smartphone based malware [19]. Different from Android malware, the Conficker worm is an Internet based state-of-the-art botnet [20]. Both the data sets have been widely used by the community.From the Android malware data set, we have an overview of the malware development from August 2010 to October 2011. There are 1,260 samples in total from 49 different Android malware in the data set. For a given Android malware program, it only focuses on one or a number of specific vulnerabilities.Therefore, all smartphones share these vulnerabilitiesform a specific network for that Android malware. In other words, there are 49 networks in the data set,and it is reasonable that the population of each network is huge.

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

FURTHER DISCUSSION:

In this paper, we have explored the problem of malware distribution in large-scale networks. There are many directions that could be further explored. We list some important ones as follows.

1) The dynamics of the late stage. We have found that the main body of malware distribution follows the power law with a short exponential tail at the late stage. It is very attractive to explore the mathematical mechanism of how the propagation leads to such kinds of mixed distributions.

2) The transition from exponential distribution to power law distribution. It is necessary to investigate when and how a malware distribution moves from
an exponential distribution to the power law. In other words, how can we clearly define the transition point between the early stage and the late stage.

3) Multiple layer modelling. We hire the fluid model in both of the two layers in our study as both layers are sufficiently large and meet the conditions for the modelling methods.

4) Epidemic model for the proposed two layer model. In this paper, we use the SI model, which is the simplest for epidemic analysis. More practical models, e.g., SIS or SIR, could be chosen to serve the same problem.

5) Distribution of coexist multiple malware in networks. In reality, multiple malware may coexist at the same networks. Due to the fact that different malware focus on different vulnerabilities, the distributions of different malware should not be the same. It is challenging and interesting to establish mathematical models for multiple malware distribution in terms of networks.

CONCLUSIONS:

In this paper, we thoroughly explore the problem of malware distribution at large-scale networks. The solution to this problem is desperately desired by cyber defenders as the network security community does not yet have solid answers. Different from previous modelling methods, we propose a two layer epidemic model: the upper layer focuses on networks of a large scale networks, for example, domains of the Internet; the lower layer focuses on the hosts of a given network. This two layer model improves the accuracy compared with the available single layer epidemic models in malware modelling. Moreover, the proposed two layer model offers us the distribution of malware in terms of the low layer networks.We perform a restricted analysis based on the proposed model, and obtain three conclusions: The distribution for a given malware in terms of networks follows exponential distribution, power law distribution with a short exponential tail, and power law distribution, at its early, late, and final stage, respectively. In order to examine our theoretical findings, we have conducted extensive experiments based on two real-world large-scale malware, and the results confirm our theoretical claims.

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] B. Stone-Gross, M. Cova, L. Cavallaro, B. Gilbert, M. Szydlowski, R. Kemmerer, C. Kruegel, and G. Vigna, "Your botnet is my botnet: Analysis of a botnet takeover," in Proc. ACM Conf. Comput. Commun. Security, 2009, pp. 635–647.

[2] M. A. Rajab, J. Zarfoss, F. Monrose, and A. Terzis, "My botnet is bigger than yours (maybe, better than yours): Why size estimates remain challenging," in Proc. 1st Conf. 1st Workshop Hot Topics Understanding Botnets, 2007, p. 5.

[3] D. Dagon, C. Zou, and W. Lee, "Modeling botnet propagation using time zones," in Proc. 13th Netw. Distrib. Syst. Security Symp., 2006.

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

[4] 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.

[5] Cabir. (2014). [Online]. Available: http://www.f-secure.com/en/ web/labs_global/2004-threat-summary.

[6] Ikee. (2014). [Online]. Available: http://www.f-secure.com/vdescs/worm_iphoneos_ikee_b.shtml

[7] Brador. (2014). [Online]. Available: http://www.f-secure.com/vdescs/brador.shtml

[8] S. Peng, S. Yu, and A. Yang, "Smartphone malware and its propagation modeling: A survey," IEEE Commun. Surveys Tuts., vol. 16, no. 2, pp. 925–941, 2014.

[9] Z. Chen and C. Ji, "An information-theoretic view of networkaware malware attacks," IEEE Trans. Inf. Forensics Security, vol. 4, no. 3, pp. 530–541, Sep. 2009.

[10] A. M. Jeffrey, X. Xia, and I. K. Craig, "When to initiate HIV therapy: A control theoretic approach," IEEE Trans. Biomed. Eng., vol. 50, no. 11, pp. 1213–1220, Nov. 2003.

[11] R. Dantu, J. W. Cangussu, and S. Patwardhan, "Fast worm containmentusing feedback control," IEEE Trans. Dependable Secure Comput., vol. 4, no. 2, pp. 119–136, Apr.–Jun. 2007.

[12] 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.

Author's:

T.Rambabu is a student of Sree Rama Institute of Technology & Science, Kuppenakuntla, Penuballi, Khammam, TS,India.Presently he is Pursuing his M.Tech (CSE) from this college His area of interests includes Information Security, Cloud Computing, Data Communication & Networks.

Mr. N.Naveen is an efficient teacher, received M.Tech from JNTU Hyderabad is working as an Assistant Professor in Department of C.S.E, Sree Rama Institute of Technology & Science, Kuppenakuntla, Penuballi, Khammam, AP,India. He has published many papers in both National & International Journals. His area of Interest includes Data Communications & Networks, Database Management Systems, Computer Organization, C Programming and other advances in Computer Applications.