MEASURING MALWARE DISTRIBUTION IN LARGE SCALE NETWORKS

R.KALAI VANI[1], Miss P.NITHYA[2]

[1]M.Phil, Computer Science, [2] M.C.A., M.Phil, Bachelor of Computer Application,
Nadar Saraswathi College of Arts and Science, (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 propagate 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.

Keywords: Malware, Propagation, Modeling, Power law, Epidemic Model

I.INTRODUCTION

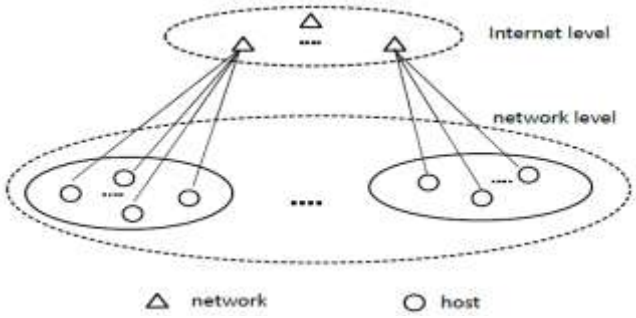
A network is a group of two or more computer systems linked together. There are many types of computer networks, including .A computer network or data network is a telecommunications network which allows computers to exchange data. In computer networks, networked computing devices pass data to each other along data connections (network links). Data is transferred in the form of packets. The connections between nodes are established using either cable media or wireless media. The best-known computer network is the Internet

Network computer devices that originate, route and terminate the data are called network nodes. Nodes can include hosts such as personal computers, phones, servers as well as networking hardware. Two such devices are said to be networked together when one device is able to exchange information with the other device, whether or not they have a direct connection to each other.

Computer networks differ in the transmission media used to carry their signals, the communications protocols to organize network traffic, the network's size, topology and organizational intent. In most cases, communications protocols are layered on (i.e. work using) other more specific or more general communications protocols, except for the physical layer that directly deals with the transmission media. Computer networks support applications such as access to the World Wide Web, shared use of application and storage servers, printers, and fax machines, and use of email and instant messaging applications.

II. COMPLEX NETWORK

A complex network is a graph (network) with non-trivial topological features that do not occur in simple networks such as lattices or random graphs but often occur in graphs modeling real systems. The study of complex networks is a young and active area of scientific research inspired largely by the empirical study of real-world networks such as computer networks and social networks.



The 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, 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. The use of Zipf distributions to represent the power law. The Zipf expression is as follows

$$\Pr\{x = i\} = C / i^{a}$$

where C is a constant, _ is a positive parameter, called the Zipf index, Pr{x=i} represents the probability of the ith (i = 1,2….) largest object in terms of size, and $\sum_i \Pr\{ x = i \} = 1$.

A more general form of the distribution is called the Zipf-Mandelbrot distribution which is defined,

$$\Pr\{ x = i\} = C / (I + q)^{\alpha}$$

where the additional constant q (q >= 0) is called the plateau factor, which makes the probability of the highest ranked objects flat. The Zipf- Mandelbrot distribution becomes the Zipf distribution when q = 0. Currently, the metric to say a distribution is a power law is to take the loglog plot of the data, and we usually say it is a power law if the result shows a straight line. We have to note that this is not a rigorous method, however, it is widely applied in practice. Power law distributions enjoy one important property, scale free. We refer interested readers to about the power law and its properties

III. MALWARE

Malware, short for malicious software, is any software used to disrupt computer operation, gather sensitive information, or gain access to private computer systems. Malware is defined by its malicious intent, acting against the requirements of the computer user, and does not include software that causes unintentional harm due to some deficiency. The term barware is sometimes used, and applied to both true (malicious) malware and unintentionally harmful software.

Malware may be stealthy, intended to steal information or spy on computer users for an extended period without their knowledge, as for example Reign, or it may be designed to cause harm, often as sabotage (e.g., Stunt), or to extort payment (CryptoLocker). 'Malware' is an umbrella term used to refer to a variety of forms of hostile or intrusive software,[1]including computer viruses, worms, trojanhorses, ransomware, spyware, adware, scareware , and other malicious programs. It can take the form of executable code, scripts, active content, and other software. Malware is often disguised as, or embedded in, non-malicious files.

As of 2011 the majority of active malware threats were worms or trojans rather than viruses. In law, malware is sometimes known as a computer contaminant, as in the legal codes of several U.S. states.

Spyware or other malware is sometimes found embedded in programs supplied officially by companies, e.g., downloadable from websites, that appear useful or attractive, but may have, for example, additional hidden tracking functionality that gathers marketing statistics. An example of such software, which was described as illegitimate, is the Sony rootkit, a Trojan embedded into CDs sold by Sony, which silently installed and concealed itself on purchasers' computers with the intention of preventing illicit copying; it also reported on users' listening habits, and unintentionally created vulnerabilities that were exploited by unrelated malware.

IV. VULNERABILITY TO SCANNER

Malware exploits security defects (security bugs or vulnerabilities) in the design of the operating system, in applications (such as browsers, e.g. older versions of Microsoft Internet Explorer supported by Windows XP), or in vulnerable versions of browser plugins such as Adobe Flash Player, Adobe Acrobat or Reader, or Java (see Java SE critical security issues). Sometimes even installing new versions of such plugins does not automatically uninstall old versions. Security advisories from plug-in providers announce security-related updates. Common vulnerabilities are assigned CVE IDs and listed in the US National Vulnerability Database. Secunia PSI is an example of software, free for personal use, that will check a PC for vulnerable out-of-date software, and attempt to update it.

Malware authors target bugs, or loopholes, to exploit. A common method is exploitation of a buffer overrun vulnerability, where software designed to store data in a specified region of memory does not prevent more data than the buffer can accommodate being supplied. Malware may provide data that overflows the buffer, with malicious executable code or data after the end; when this payload is accessed it does what the attacker, not the legitimate software, determines. Early PCs had to be booted from floppy disks; when built-in hard drives became common the operating system was normally started from them, but it was possible to boot from another boot device if available, such as a floppy disk, CD-ROM, DVD-ROM, or USB flash drive. It was common to configure the computer to boot from one of these devices when available. Normally none would be available; the user would intentionally insert, say, a CD into the optical drive to boot the computer in some special way, for example to install an operating system. Even without booting, computers can be configured to

execute software on some media as soon as they become available, e.g. to auto run a CD or USB device when inserted.

V. BOTNET

A botnet is a collection of Internet-connected programs communicating with other similar programs in order to perform tasks. This can be as mundane as keeping control of an Internet Relay Chat (IRC) channel, or it could be used to send spam email or participate in distributed denial-of-service attacks. The word botnet is a combination of the wordsrobot and network. The term is usually used with a negative or malicious connotation

5.1. Legal Botnets

The term botnet is widely used when several IRC bots have been linked and may possibly set channel modes on other bots and users while keeping IRC channels free from unwanted users. This is where the term is originally from, since the first illegal botnets were similar to legal botnets. A common bot used to set up botnets on IRC is eggdrop.

5.2. Illegal Botnets

Botnets sometimes compromise computers whose security defenses have been breached and control conceded to a third party. Each such compromised device, known as a "bot", is created when a computer is penetrated by software from a *malware* (malicious software) distribution. The controller of a botnet is able to direct the activities of these compromised computers through communication channels formed by standards-based network protocols such as IRC and Hypertext Transfer Protocol (HTTP).

VI. MEASURE THE SIZE OF BOTNETS

The prolific Flashback botnet of Macintosh computers on one day last month was counted at anywhere from more than half a million to 1 million bots worldwide. One security firm later reported infections dropping to tens of thousands, while another found 700,000 bots still phoning home to the botnet operator infrastructure. Yet another says the total number of infected Macs was even higher than was originally reported.

6.1 Botnet Infiltration

An obvious way to learn several aspects of a botnet's activity is to infiltrate the botnet by joining the command and control channel. Botnet infiltration provides valuable information about several malicious activities such as DDoS attacks as shown earlier by Freiling . In our earlier work , we used botnet infiltration to provide in-depth analysis of several facets of botnets, including inferring their membership by directly counting the bots observed on individual command and control channels.

6.2 DNS Redirection

DNS hijacking or DNS redirection is the practice of subverting the resolution of Domain Name System (DNS) queries. This can be achieved by malware that overrides a computer's TCP/IP configuration to point at a rogue DNS server under the control of an attacker, or through modifying the behaviour of a trusted DNS server so that it does not comply with internet standards.

These modifications may be made for malicious purposes such as phishing, or for self-serving purposes by Internet service providers (ISPs) and public/router-based online DNS server providers to direct users' web traffic to the ISP's own web servers where advertisements can be served, statistics collected, or other purposes of the ISP; and by DNS service providers to block access to selected domains as a form of censorship

VII. LITERATURE SURVEY

"Modeling malware propagation in Gnutella type peer to peer networks Parallel and Distributed" …, K Ramachandran, B Sikdar…, Volume pp no 2.., year 2006

A key emerging and popular communication model mostly invented for getting information which is peer-to-peer (P2P) networking. To spread of malware in decentralized Gnutella type of peer-to-peer network is needed. The study reveals that the existing bound on the spectral radius governing the possibility of an epidemic outbreak needs to be revised in the context of a P2P network. To formulate an analytical model that reveals the study of mechanics and decentralized Gnutella type of peer network and study the spread of malware on such networks. The show analytically, that a framework which does not incorporate the behavioral characteristics of peers. This in turn results in negatives, an undesirable feature. Thus differentiating the conditions under which the network may reach a malware free equilibrium and validate the theoretical results with numerical simulations

" A Large Scale Empirical study of conficker" S Shin, G Gu, N Reddy, CP Lee.., volume pp no 3…, IEEE year 2012.

If 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 Conflicker, the 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. On observing that the Conflicker 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.If measure the potential power of Conficker to estimate its effects on the networks when it performs malicious operations.

"Smartphone Malware and its propagation Modeling's survey" S Peng , S Yu, A Yang …, volume pp no 3…, IEEE year 2013

Smartphones are used in society, and have been both the target and sufer of malware writers. Motivated by the significant threat that presents for proper using of users, by survey the current smart phone malware status and their propagation models. The content of this paper is presented in two parts. In the first part, we review the short history of mobile malware evolution since 2004, and then list the classes of mobile malware and their infection vectors. At the end of the first part, explain the possible damage caused by smart phone malware. The second part, focuses on smart phone malware propagation modeling. In order to understand the propagation behavior of smart phone malware, recall generic epidemic models as a foundation for further exploration, then extensively survey the smart phone malware propagation models. At the end of this paper, the highlight issues of the current smart phone malware propagation models and discuss about possible future trends based on our understanding of this topic.

VIII. CONCLUSION

The solution for problem is desperately desired by cyber defenders as the network security community does not yet have solid answers. It is different from previous modeling methods, as such as two layer epidemic model has deviced. It is upper layer focuses on networks of a large scale networks, 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.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,to have conducted extensive experiments based on two real-world large-scale malware, and the results confirm our theoretical claims.

IX. FUTURE ENHANCEMENT

In regards to future work, firstly 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. Secondly, defenders may care more about their own network, the distribution of a given malware at their ISP domains, where the conditions for the two layer model may not hold. Need to seek appropriate models to address this problem. Finally, to studying the distribution of multiple malware on large-scale networks such as only focus on one malware in this paper. 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 CCS ’09: Proceedings of the 2009 ACM conference on computer communication security, 2009.

[2]. D. Dagon, C. Zou, andW. Lee, “Modeling botnet propagation using time zones,” in Proceedings of the 13 th Network and Distributed System Security Symposium NDSS, 2006.

[3]. 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 Proceedings of the first conference on First Workshop on Hot Topics in Understanding Botnets, 2007.

[4]. D. Dagon, C. C. Zou, and W. Lee, “Modeling botnet propagation using time zones,” in NDSS, 2006.

[5]. P. V. Mieghem, J. Omic, and R. Kooij, “Virus spread in networks,” IEEE/ACM Transactions on Networking, vol. 17, no. 1, pp. 1–14, 2009.

[6]. Cabir, http://www.f-secure.com/en/web/labs global/2004- threat-summary.

[7]. S. Peng, S. Yu, and A. Yang, “Smartphone malware and its propagation modeling: A survey,” IEEE Communications Surveys and Tutorials, in press, 2014.

[8]. Z. Chen and C. Ji, “An information-theoretic view of network-aware malware attacks,” IEEE Transactions on Information Forensics and Security, vol. 4, no. 3, pp. 530– 541, 2009.

[9]. A.M. Jeffrey, xiaohua Xia, and I. K. Craig, “When to initiate hiv therapy: A control theoretic approach,” IEEE Transactions on Biomedical Engineering, vol. 50, no. 11, pp. 1213–1220, 2003.

[10]. R. Dantu, J.W. Cangussu, and S. Patwardhan, "Fast worm containment using feedback control," IEEE Transactions on Dependable and Secure Computing, vol. 4, no. 2, pp. 119–136, 2007.

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

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

[13]. G. Yan and S. Eidenbenz, "Modeling propagation dynamics of bluetooth worms (extended version)," IEEE Trans. Mob. Comput., vol. 8, no. 3, pp. 353–368, 2009.