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. 3, Issue 11, November 2015

A Review on: Malware Propagation in Large-Scale Networks

Dhende Kapil N., Prof. Bere S. S.

P.G. Scholar, Dept. of Computer Engineering. DGOI, FOE, Bhigwan, Savitribai Phule University of Pune, Pune, India

Professor, Dept. of Computer Engineering, DGOI, FOE, Bhigwan, Savitribai Phule University of Pune, Pune, 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.

KEYWORDS*:* Malware, propagation, modelling, power law.

I.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.

II.LITERATURE SURVEY

1**. "Modeling botnet propagation using time zones,".**

AUTHORS: D. Dagon, C. Zou, andW. Lee
Time zones play an important and unexplored role in malware epidemics. To understand how time and location affect malware spread dynamics, we studied botnets, or large coordinated collections of victim machines (zombies) controlled by attackers. Over a six month period we observed dozens of botnets representing millions of victims. We noted diurnal properties in botnet activity, which we suspect occurs because victims turn their computers off at night. Through binary analysis, we also confirmed that some botnets demonstrated a bias in infecting regional populations. Clearly, computers that are offline are not infectious, and any regional bias in infections will affect the overall growth of the botnet. We therefore created a diurnal propagation model. The model uses diurnal shaping functions to capture regional variations in online vulnerable populations. The diurnal model also lets one compare propagation rates for different botnets, and prioritize response. Because of variations in release times and diurnal shaping functions particular to an infection, botnets released later in time may actually surpass other botnets that have an advanced start. Since response times for malware outbreaks is now measured in hours, being able to predict short-term propagation dynamics lets us allocate resources more intelligently. We used empirical data from botnets to evaluate the analytical model.

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. 3, Issue 11, November 2015

2. **"Dissecting android malware: Characterization and evolution,".**
AUTHORS: Y. Zhou and X. Jiang
The popularity and adoption of smart phones has greatly stimulated the spread of mobile malware, especially on the popular platforms such as Android. In light of their rapid growth, there is a pressing need to develop effective solutions. However, our defense capability is largely constrained by the limited understanding of these emerging mobile malware and the lack of timely access to related samples. In this paper, we focus on the Android platform and aim to systematize or characterize existing Android malware. Particularly, with more than one year effort, we have managed to collect more than 1,200 malware samples that cover the majority of existing Android malware families, ranging from their debut in August 2010 to recent ones in October 2011. In addition, we systematically characterize them from various aspects, including their installation methods, activation mechanisms as well as the nature of carried malicious payloads. The characterization and a subsequent evolution-based study of representative families reveal that they are evolving rapidly to circumvent the detection from existing mobile anti-virus software. Based on the evaluation with four representative mobile security software, our experiments show that the best case detects 79.6% of them while the worst case detects only 20.2% in our dataset. These results clearly call for the need to better develop next-generation anti-mobile-malware solutions.

3. **"Protecting against network infections: A game theoretic perspective,".**
AUTHORS: J. Omic, A. Orda, and P. V. Mieghem
Security breaches and attacks are critical problems in today's networking. A key-point is that the security of each host depends not only on the protection strategies it chooses to adopt but also on those chosen by other hosts in the network. The spread of Internet worms and viruses is only one example. This class of problems has two aspects. First, it deals with epidemic processes, and as such calls for the employment of epidemic theory. Second, the distributed and autonomous nature of decision-making in major classes of networks (e.g., P2P, ad- hoc, and most notably the Internet) call for the employment of game theoretical approaches. Accordingly, we propose a unified framework that combines the N-intertwined, SIS epidemic model with a no cooperative game model. We determine the existence of Nash equilibrium of the respective game and characterize its properties. We show that its quality, in terms of overall network security, largely depends on the underlying topology. We then provide a bound on the level of system inefficiency due to the no cooperative behaviour, namely, the "price of anarchy" of the game. We observe that the price of anarchy may be prohibitively high; hence we propose a scheme for steering users towards socially efficient behaviour.

4. **"Power laws, pareto distributions and zipf's law,".**
AUTHORS: M. E. J. Newman,
When the probability of measuring a particular value of some quantity varies inversely as a power of that value, the quantity is said to follow a power law, also known variously as Zipf's law or the Pareto distribution. Power laws appear widely in physics, biology, earth and planetary sciences, economics and finance, computer science, demography and the social sciences. For instance, the distributions of the sizes of cities, earthquakes, solar flares, moon craters, wars and people's personal fortunes all appear to follow power laws. The origin of power-law behavior has been a topic of debate in the scientific community for more than a century. Here we review some of the empirical evidence for the existence of power-law forms and the theories proposed to explain them.

5. **"The effect of network topology on the spread of epidemics,".**
AUTHORS: A. J. Ganesh, L. Massouli´e, and D. F. Towsley
Many network phenomena are well modeled as spreads of epidemics through a network. Prominent examples include the spread of worms and email viruses, and, more generally, faults. Many types of information dissemination can also be modeled as spreads of epidemics. In this paper we address the question of what makes an epidemic either weak or potent. More precisely, we identify topological properties of the graph that determine the persistence of epidemics. In particular, we show that if the ratio of cure to infection rates is larger than the spectral radius of the graph, then the mean epidemic lifetime is of order log n, where n is the number of nodes. Conversely, if this ratio is smaller than a generalization of the isoperimetric constant of the graph, then the mean epidemic lifetime is of order ena, for a positive constant a. We apply these results to several network topologies including the hypercube, which is a representative connectivity graph for a distributed hash table, the complete graph, which is an important connectivity graph for BGP,

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. 3, Issue 11, November 2015

and the power law graph, of which the AS-level Internet graph is a prime example. We also study the star topology and the Erdos-Renyi graph as their epidemic spreading behaviors determine the spreading behavior of power law graph.

Existing System:

The epidemic theory plays a leading role in malware propagation modelling. The current models for malware spread fall in two categories: the epidemiology model and the control theoretic model. The control system theory based models try to detect and contain the spread of malware. The epidemiology models are more focused on the number of compromised hosts and their distributions, and they have been explored extensively in the computer science community. Zou et al. used a susceptible-infected (SI) model to predict the growth of Internet worms at the early stage. Gao and Liu recently employed a susceptible-infected-recovered (SIR) model to describe mobile virus propagation.

Disadvantages:

- ✓ One critical condition for the epidemic models is a large vulnerable population because their principle is based on differential equations.
- ✓ As pointed by Willinger et al. the findings, which we extract from a set of observed data, usually reflect parts of the studied objects. It is more reliable to extract the-oretical results from appropriate models with confirmation from sufficient real world data set experiments.

III. PROPOSED ALGORITHM

System Architecture:

In this paper, we study the distribution of malware in terms of networks (e.g., autonomous systems, ISP domains, and abstract net-works 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 breakout of a malware, we calculate how many networks have been compromised based on the SI model. Secondly, for a compromised net-work, we calculate how many hosts have been compromised since the time that the network was compromised.

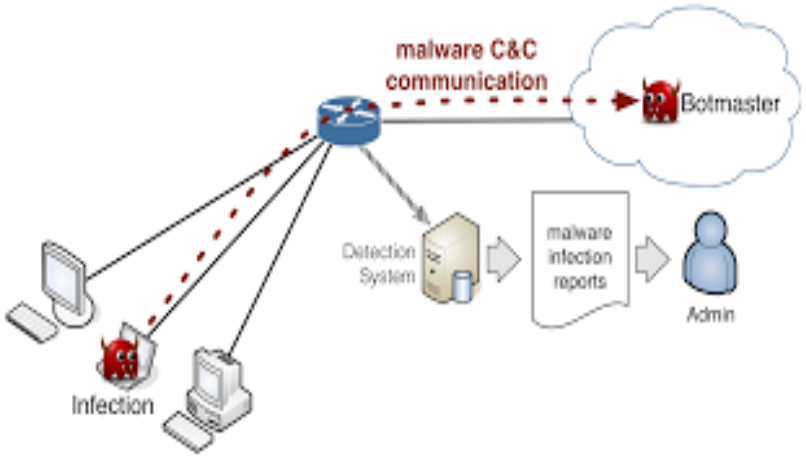

Figure 1: System Architecture of Proposed System.

Modules:

1. **Network Formation:**
   Research on complex networks has 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

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. 3, Issue 11, November 2015

2. Malware Propagation:

a) Early stage: An early stage of the breakout of a malware means only a small percentage of vulnerable hosts have been compromised, and the propagation follows exponential distributions.

b) Final stage: The final stage of the propagation of a malware means that all vulnerable hosts of a given network have been compromised.

c) Late stage: A late stage means the time interval between the early stage and the final stage.

3. Filtering Malware Detection

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.The two layers in both layers are sufficiently large and meet the conditions for the modelling methods. In order to improve the accuracy of malware propagation, we may extend our work to layers. In another scenario, we may expect to model a malware distribution for middle size networks

4. Performance Evaluation

We have to note that our experiments also indicate that this data does not fit the power law. For a given Android malware program, it only focuses on one or a number of specific vulnerabilities. Therefore, all smartphones share these vulnerabilities form a specific network for that Android malware.

Advantage:

a. 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.

IV. CONCLUSION AND FUTURE WORK

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 modeling 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.

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] Ikee, http://www.f-secure.com/vdescs/worm iphoneosikee b.shtml.
[8] Brador, http://www.f-secure.com/v-descs/brador.shtml.

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. 3, Issue 11, November 2015

[9] S. Peng, S. Yu, and A. Yang, "Smartphone malware and its propagation modeling: A survey," IEEE Communications Surveys and Tutorials, in press, 2014.
[10] 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.
[11] 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.
[12] 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.
[13] 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.
[14] P. De, Y. Liu, and S. K. Das, "An epidemic theoretic framework for vulnerability analysis of broadcast protocols in wireless sensor networks," IEEE Trans. Mob. Comput., vol. 8, no. 3, pp. 413–425, 2009.
[15] 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.
[16] C. C. Zou, W. Gong, D. Towsley, and L. Gao, "The monitoring and early detection of internet worms," IEEE/ACM Trans. Netw., vol. 13, no. 5, pp. 961–974, 2005.
[17] C. Gao and J. Liu, "Modeling and restraining mobile virus propagation," IEEE Trans. Mob. Comput., vol. 12, no. 3, pp. 529–541, 2013.
[18] D. J. Daley and J. Gani, Epidemic Modelling: An Introduction. Cambridge University, 1999.
[19] W. Willinger, D. Alderson, and J. C. Doyle, "Mathematics and the internet: A source of enormous confusion and great potential," Notices of the Ameriacan Mathematical Socieity, vol. 56, no. 5, pp. 586–599, 2009.
[20] Y. Zhou and X. Jiang, "Dissecting android malware: Characterization and evolution," in IEEE Symposium on Security and Privacy, 2012, pp. 95–109.
[21] S. Shin, G. Gu, A. L. N. Reddy, and C. P. Lee, "A largescale empirical study of conficker," IEEE Transactions on Information Forensics and Security, vol. 7, no. 2, pp. 676–690, 2012.
[22] M. A. Rajab, J. Zarfoss, F. Monrose, and A. Terzis, "A multifaceted approach to understanding the botnet phenomenon," in Internet Measurement Conference, 2006, pp. 41–52.
[23] A. J. Ganesh, L. Massouli´e, and D. F. Towsley, "The effect of network topology on the spread of epidemics," in INFOCOM, 2005, pp. 1455–1466.
[24] J. Omic, A. Orda, and P. V. Mieghem, "Protecting against network infections: A game theoretic perspective," in INFOCOM'09, 2009.
[25] R. L. Axtell, "Zipf distribution of u.s. firm sizes," Science, vol. 293, 2001.
[26] M. Mitzenmacher, "A brief history of generative models for power law and lognornal distributions," Internet Mathematics, vol. 1, 2004.
[27] M. Newman, Networks, An Introduction. Oxford University Press, 2010.
[28] Z. K. Silagadze, "Citations and the zipf-mandelbrot's law," Complex Systems, vol. 11, pp. 487–499, 1997.
[29] M. E. J. Newman, "Power laws, pareto distributions and zipf's law," Contemporary Physics, vol. 46, pp. 323–351, December 2005.
[30] L. Kleinrock, Queueing Systems. Wiley Interscience, 1975, vol. I: Theory.