A Survey on Malware Propagation in Large Scale Network

MISS. ANKITA SURESH MANE[1], PROF. SAYYAD G.G.[2]

*[1] PG Scholar, Department of Computer Engineering Dattakala Faculty of Engineering
Pune, Maharashtra, India*

*[2] Professor, Department of Computer Engineering Dattakala Faculty of Engineering
Pune, Maharashtra, India*

---***---

Abstract - ***Malware is prevalent in networks, and poses a critical threat to network security. However, we not fully known of malware behavior in networks to date. In this paper, we examine from a global perspective, how malware propagates in networks. We formulate the problem, and establish a rigorous two layer epidemic model from network to network for malware propagation. Based on the proposed model, our analysis show 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 stage, late stage 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.*

1.INTRODUCTION *(Size 11 , cambria font)*

A network is a group of two or more computer systems linked together. A computer network or data network is a telecommunications network. It allows computers to exchange data. In computer networks devices transfer data to each other along data connections. The most-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 and also networking hardware. 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.

Malware is software used to disrupt computer operation, gather sensitive information, or gain access to private computer systems. A compromised computer is known as bot, is created when a computer is penetrated by software from a malware (malicious software) distribution. Botnets sometimes compromise computers whose security defenses have been breached and control conceded to a third party. Botnets have become the attack engine of cyber attackers, and they pose critical challenges to cyber defenders. it is important for defenders to understand malware behavior, in order to fight against cyber criminals.

2. LITERATURE SUREY

A. Method: 1

"Modeling botnet propagation using time zones," D. Dagon, C. Zou , and W. Lee[1] : Introduced time zone information, and built a model to describe the impact on the number of live members of botnets with diurnal effect. Time zones play an important role in malware epidemics. We studied botnets to understand how time and location affect malware spread dynamics. We observed dozens of botnets representing millions of victims, over a six month period. Because victims turn their computers off at night, in botnet activity we noted diurnal properties, which we suspect occurs. We also confirmed that some botnets demonstrated a bias in infecting regional populations, by the binary analysis. Computers are not contagious that are offline and any regional bias will affect the overall growth of the bone, in contagious for that purpose we create a diurnal propagation model. To capture regional variations in online vulnerable populations the model uses diurnal shaping functions. This model also compare propagation rates for different botnets, and prioritize response. Botnets released later in time may actually surpass other botnets that have an advanced start because of variations in release times and diurnal shaping functions particular to an infection,. Since response times for malware outbreaks is measured in hours, being able to estimate short-term propagation dynamics lets us allocate resources more intelligently. We used empirical data from botnets to evaluate the analytical model.

B. Method :2

,"A large-scale empirical study of conficker S. Shin, G. Gu, A. L. N. Reddy, and C. P. Lee [6] : Conficker is a known and most widespread malware. collected a data set about 25 million victims, and study various interesting aspects about this state-of-the-art malware. By analysing Conflicker, the main intention to understand current and new trends in malware propagation, which could be very useful in estimating future malware trends and providing insights for future malware defense. On observing that the Conflicker has some different victim distribution patterns that compared to many previous generation worms, 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.

C. Method : 3

"Smartphone malware and its propagation modeling A survey," S Peng , S Yu, A Yang[3]: presented the short history of mobile malware since 2004, and surveyed their propagation models. Smartphones are commonly used in society, and have been both the target and victim 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 mobile malware classes 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 propagation modelling of smart phone malware. In order to understand the propagation behaviour of smart phone malware, recall generic epidemic models as a foundation for further exploration, then survey the smart phone malware propagation models. At the end of this paper, the current issues of the smart phone malware propagation models and discuss the possible future trends based on our understanding of this topic.

D. Method : 4

"Power laws, pareto distributions and zipf's law," In 2005 M. E. J. Newman[4]: Introduced power law. The functional relationship between two quantities is power law. 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. Power law is also known as Zipf's law or the Pareto distribution. Power laws mostly come in physics, biology, economics and finance, computer science, demography and the social sciences. For instance, the distributions of the sizes of cities, earthquakes, solar flares, moon craters, people's personal fortunes all appear to follow power laws. Here we review some of the empirical evidence for the existence of power-law forms and the theories proposed to explain them.

3. EXISTING SYSTEM

All paragraphs must be indented. In malware propagation modeling the epidemic theory plays a main role. There are two categories in current models for malware spread :the epidemiology model and the control theoretic model. The control theoretic models try to detect and contain the spread of malware. The epidemiology model focused on the number of compromised hosts and their distributions, and they have been travel extensively in the computer science community. To predict the growth of Internet worms, Zou et al. used a susceptible-infected (SI) model at the early stage. To describe mobile virus propagation, Gao and Liu recently employed a susceptible-infected-recovered (SIR) model.

Disadvantages:

- One critical condition for the epidemic models is a large vulnerable population because their principle is depends 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.

4. PROPOSED SYSTEM

In these paper, at large scales we study the distribution of malware in terms of networks. In this setting, to meet the requirements of the SI model, we have a sufficient volume of data at a large enough scale. we break our model into two layers, from the traditional epidemic models. In first layer, for a given time since the breakout of a malware and we calculate based on the SI model, how many networks have been compromised. Secondly, we calculate how many hosts have been compromised since the time that the network was compromised, for compromised net-work.

A. Network Formation:

Research on complex networks has indicated 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.

B. 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.

C. Filtering Malware Detection:

In reality, multiple malware may in the same place 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, in terms of networks to establish mathematical models for multiple malware distribution. The two layers in both layers are sufficiently large and meet the conditions for the modelling methods. We may extend our work to layers in order to improve the accuracy of malware propagation.

D. Performance Evaluation

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

Advantage:

a. Our rigorous analysis, at its early stage, we find that the distribution of a given malware follows an exponential distribution and at its late stage, obeys a power law distribution with a short exponential tail and finally converges to a power law distribution.

5. CONCLUSIONS

In this paper, we explore the problem of malware distribution at large-area networks. By cyber defenders the solution to this problem is desperately desired, as the network security community does not yet have solid answers. We propose a two layer epidemic model: the upper layer focuses on networks of a large scale networks. The lower layer focuses on a given network host. In malware modeling this two layer model improves the accuracy compared with the available single layer epidemic models. Moreover, in terms of the low layer networks this two layer model offers us the distribution of malware. Based on the proposed model we perform a restricted analysis. In terms of networks the distribution for a given malware follows exponential distribution, power law distribution with a short exponential tail, and power law distribution, at its early stage, late stage, and final stage, respectively. We have conducted extensive experiments based on two real-world large-scale malware, and the results confirm our theoretical claims.

ACKNOWLEDGEMENT

I would like to express my gratitude towards my guide and H.O.D of Computer Engg. Department Prof. Sayyed G.G. for their support and guidance in my work.

REFERENCES

[1] 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.

[2]] J. Omic, A. Orda, and P. V. Mieghem, "Protecting against network infections: A game theoretic perspective," in INFOCOM'09, 2009.

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

[4] M. E. J. Newman, "Power laws, pareto distributions and zipf's law," Contemporary Physics, vol. 46, pp. 323–351, December 2005.

[5] 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.

[6] 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.

[7] J. Omic, A. Orda, and P. V. Mieghem, "Protecting against network infections: A game theoretic perspective," in INFOCOM'09, 2009.

[8] R. L. Axtell, "Zipf distribution of u.s. firm sizes," Science, vol. 293, 2001.

[9] M. Mitzenmacher, "A brief history of generative models for power law and lognornal distributions," Internet Mathematics, vol. 1, 2004.

[10] *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.*