ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 7, July 2017

Analysis of Malware Propagation in Large Scale Network

Ankita Suresh Mane, Prof. Bere S.S.

M.E. Student, Dept. of Computer Engineering, DGOIFOE, Bhigwan, Pune, India

Assistant Professor, Dept. of computer Engineering, DGOIFOE, Bhigwan, Pune, India

ABSTRACT: Malware is prevalent in networks, and poses a critical threat to network security. However, we are not fully known of malware behaviour 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, Modelling, Power law, Zipf distribution, Epidemic Model.

I. INTRODUCTION

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. In most cases, communications protocols are layered on 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. 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. Power law distributions enjoy one important property, scale free.

Malware is software used to disrupt computer operation, gather sensitive information, or gain access to private computer systems. Malware is an umbrella term used to refer to a variety of forms of hostile or intrusive software, including computer viruses, worms, Trojan horses, ransomware, spyware, adware, scareware, and other malicious programs. It can take the form of executable code, scripts, active content, and other software. 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 known as bot, is created when a computer is penetrated by software from a malware distribution. Botnets sometimes compromise computers whose \ security defences 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 behaviour, in order to fight against cyber criminals. There are various methods to measure the size of botnets, such as botnet infiltration, DNS redirection, and external information. Botnet infiltration provides valuable information about several malicious activities such as DDoS attacks and in depth analysis of several facets of botnets, including inferring their membership by directly counting the bots observed on individual command and control channels. To achieve this, a lightweight IRC tracker is used. Another method is to use DNS redirection in that analysed

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 7, July 2017

captured bots by honeypot, and then identified the CC server using source code reverse engineering tools. They then manipulated the DNS entry which is related to a botnets IRC server, and redirected the DNS requests to a local sinkhole. They therefore could count the number of bots in the botnet. Dagon, Zou and Lee [1] says that time zone has impact on the number of available bots. Mieghem et al[9] shows that the network topology has an important impact on malware spreading through their rigorous mathematical analysis. The emergence of mobile malware, such as Cabir, Ikee, and Brador further increases the difficulty level of understanding on how they propagate. To the best of our knowledge, the best finding about malware distribution in large scale networks is the distribution is non-uniform. All this indicates that the research in this field is in its early stage.

II. RELATED WORKS

In [1]authors 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. In [2]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 results with numerical simulations.In [3]authors 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. In [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 Zipfs law or the Pareto distribution. For instance, the distributions of the sizes of cities, earthquakes, solar flares, moon craters, and 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. In [5] authors describe 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 defence capability is largely constrained by the limited understanding of these emerging mobile malware and the lack of timely access to related samples.

III. PROPOSED SYSTEM

In this paper, at large area we study the distribution of malware in terms of networks. To meet the requirements of the SI model, we have a sufficient volume of data at a large area. We break our model into two layers, from the traditional epidemic models as shown in Fig.1. In first layer, for a given time since the breakout of a malware and we calculate how many networks have been compromised based on the SI model. Secondly, we calculate how many hosts have been compromised since the time that the network was compromised, for compromised 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.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 7, July 2017

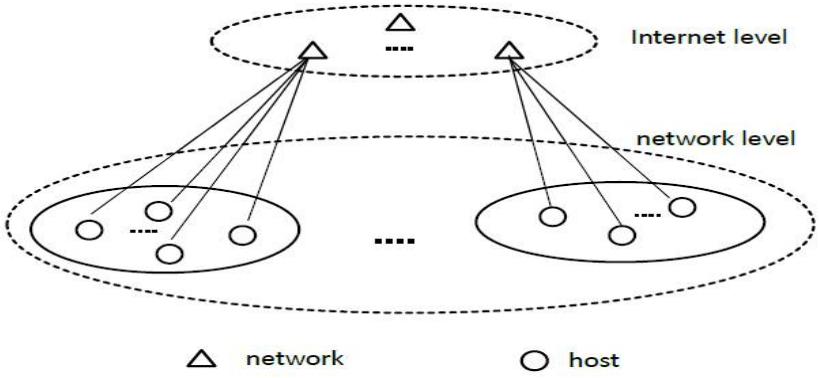

Fig.1 System Architecture

Implementation is the stage of the project when the theoretical design is turned out into a working system. In Malware propagation in large scale networks we have the modules such as discussed below

A. User

The new user or admin going to access their page then they have to register first by providing necessary details. After successful completion of sign up process, the user has to login into the application by providing username and exact password. The user has to provide exact username and password which was provided at the time of registration, if login success means it will take up to main page else it will remain in the login page itself.

B. Admin Server

The Admin server is responsible for performing some operations like to analysing documents and contents to check whether the document contains malware. If documents are malware related then those documents will be scanned and finds malicious users those who propagate malware.

C. Malware

Malware are deployed by cyber attackers to compromise computer systems by exploiting their security vulnerabilities. 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 behaviour, such as propagation or membership recruitment patterns, the size of botnets, and distribution of bots.

D. Malware Propagation

1) Early stage: In early stage only a small percentage of vulnerable hosts have been compromised.
2) Final stage: In final stage all vulnerable hosts of a given network have been compromised.
3) Late stage: the time interval between the early stage and the final stage means late stage.

E. Power Law Distribution

Complex networks have demonstrated that the number of hosts of networks follows the power law. 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. The Zipf distributions are tidier than the expression of the Pareto distributions. In this paper, we will use Zipf distributions to represent the power law. 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.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 7, July 2017

F. Performance Evaluation

Examine the theoretical analysis through well-known large-scale malware: Android malware and Conficker. The Conficker worm is an Internet based state-of-the-art botnet. Both the data sets have been widely used by the community. Sort the malware subclasses according to their size, and present them a loglog format. A unique feature of the power law is the scale free property.

Advantage:

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.

IV. RESULT ANALYSIS

We have created user login. After the user login done successfully then load the data sets and calculate probability value as shown in Fig.2.

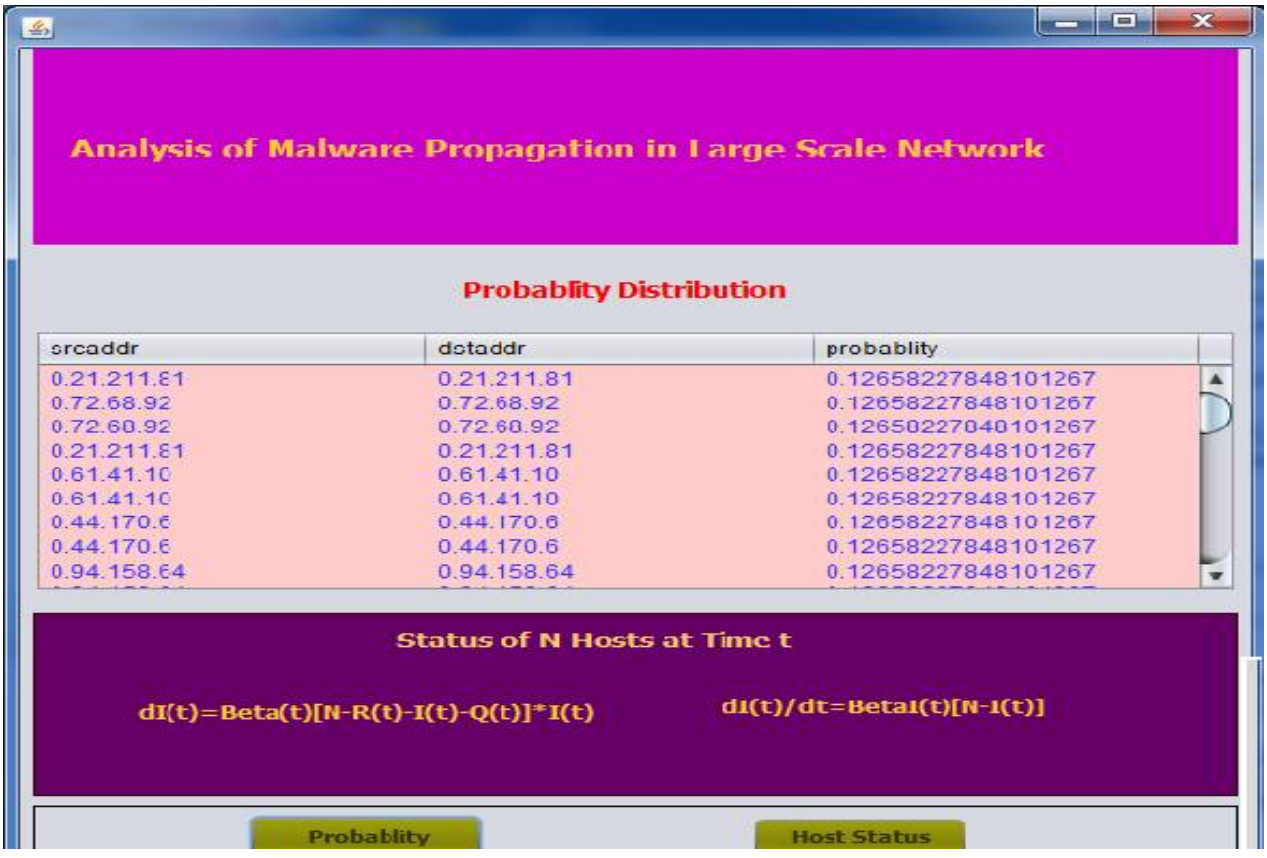

Fig.2 Probability Distribution

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 7, July 2017

From the probability distribution hosts are clearly classified in infected host and Recovered host as shown in fig.3.Then calculate infection rate.

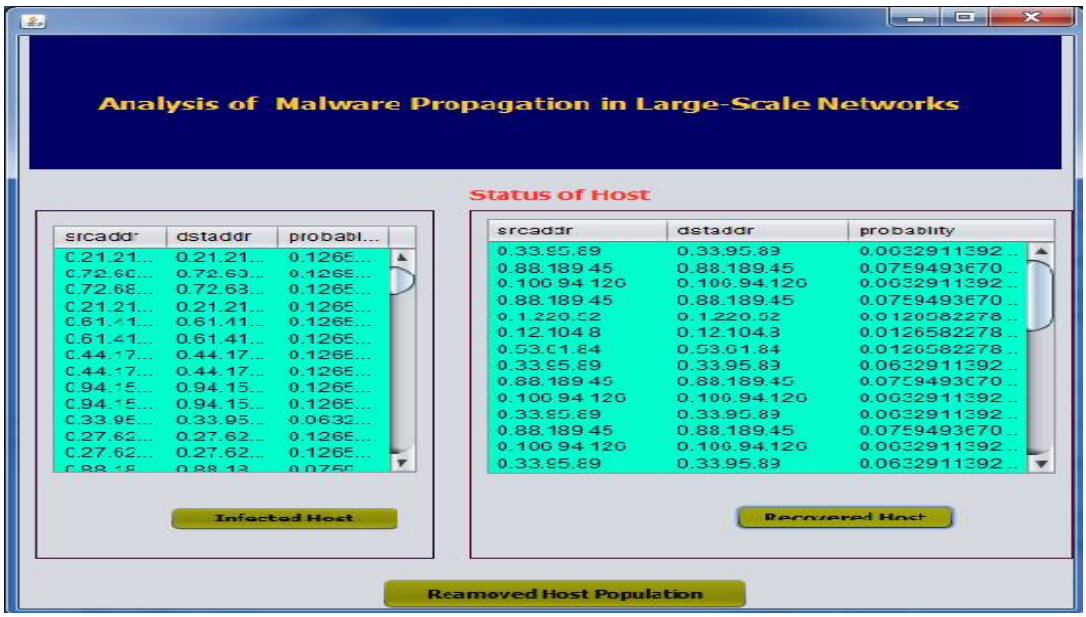

Fig. 3 Status of Hosts

Fig.4 shows the model analysis, model analysis clearly classified the hosts population in terms of initial stage, late stage and final stage.

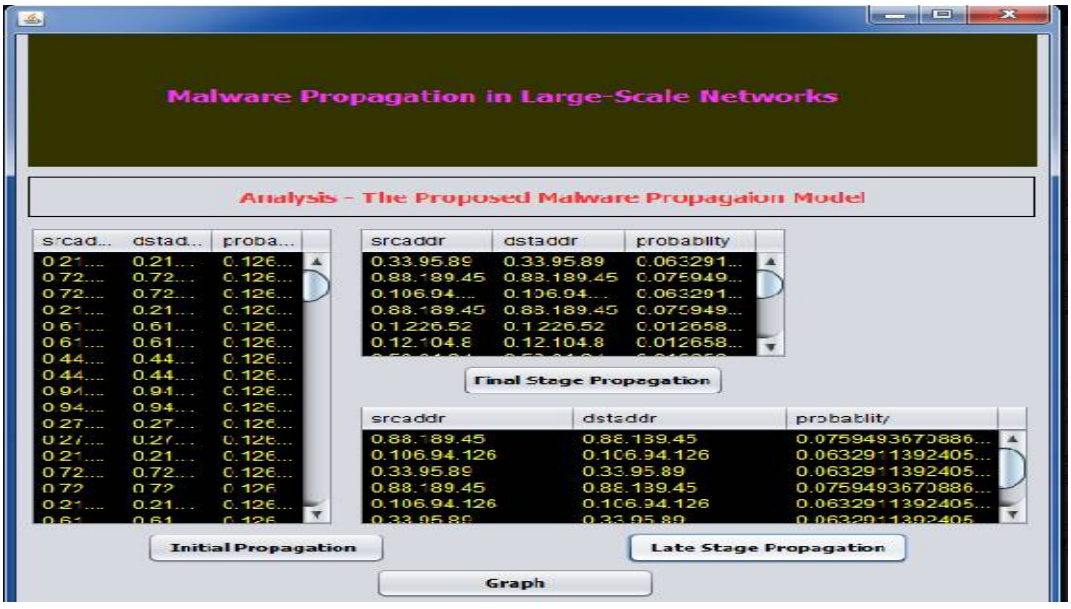

Fig.4. Model analysis of hosts

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 7, July 2017

V. CONCLUSION AND FUTURE WORK

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 modelling 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.

In this concept mainly we have to know about when and how malware distribution moves from an exponential distribution to the power law distribution. In this we focused only on one malware, In future we are interested to find distribution of multiple malware on large scale networks.

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] K. Ramachandran, B Sikdar, Modeling malware propagation in Gnutella type peer to peer network parallel distribution, Processing Symposium, 2006. IPDPS 2006.
[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 zipfs law, Contemporary Physics, vol. 46, pp. 323351, December 2005.
[5] Y. Zhou and X. Jiang, Dissecting android malware: Characterization and evolution, in IEEE Symp. Security Privacy, 2012, pp. 95109
[6] S. Shin, G. Gu, A. L. N. Reddy, and C. P. Lee, A large-scale empirical study of conficker, IEEE Transactions on Information Forensics and Security, vol. 7, no. 2, pp. 676690, 2012.
[7] A. J. Ganesh, L. Massoulie, and D. F. Towsley, The effect of network topology on the spread of epidemics, in INFOCOM, 2005, pp. 14551466.
[8] J. Omic, A. Orda, and P. V. Mieghem, Protecting against network infections: A game theoretic perspective, in INFOCOM09, 2009.
[9] 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.
[10] P. V. Mieghem, J. Omic, and R. Kooij, Virus spread in networks, IEEE/ACM Transactions on Networking, vol. 17, no. 1, pp. 114, 2009.
[11] R. L. Axtell, Zipf distribution of u.s. firm sizes, Science, vol. 293, 2001.
[12] 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.
[13] G. Yan and S. Eidenbenz, Modeling propagation dynamics of Bluetooth worms (extended version), IEEE Trans. Mob. Comput., vol.8,no. 3, pp. 353368, 2009. .