DETECTING MALWARE SPREAD

J.GOMATHI[1], S. Kayathri[2]

[1]PG Scholar, Dept of MCA, M.Kumarasamy College of Engineering, Karur, Tamil Nadu, India,
[2]Associate Professor, Dept of MCA, M.Kumarasamy College of Engineering, Karur, Tamil Nadu, India.

Abstract: Malware is pervasive in systems, and represents a basic danger to network security. Notwithstanding, we have exceptionally constrained comprehension of malware conduct in systems to date. In this paper, we explore how malware spreads in systems from a worldwide viewpoint. We figure the issue, and set up a thorough two layer plague model for malware proliferation from system to organize. Taking into account the proposed model, our examination shows that the conveyance of a given malware takes after exponential conveyance, power law dispersion with a short exponential tail, and power law circulation at its initial, late what's more, last stages, separately. Broad tests have been performed through two certifiable worldwide scale malware information sets, and the outcomes affirm our hypothetical discoveries.

Keywords: Malware, Detecting, Modeling, Power Law.

I. INTRODUCTION

Malware are noxious programming programs sent by digital assailants to bargain PC frameworks by misusing their security vulnerabilities. Spurred by unprecedented monetary or political prizes, malware proprietors are depleting their vitality to trade off the greatest number of organized PCs as they can keep in mind the end goal to accomplish their vindictive objectives. A traded off PC is known as a bot, and all bots bargained by a malware structure a botnet. Botnets have turned into the assault motor of digital assailants, and they posture basic difficulties to digital safeguards so as to battle against digital lawbreakers, it is vital for guards to comprehend malware conduct, for example, spread or enrollment enlistment designs, the span of botnets, and appropriation of bots. The pestilence hypothesis assumes a main part in malware engendering demonstrating. The current models for malware spread fall in two categories: the study of disease transmission model and the control theoretic model. The control framework hypothesis based models attempt to distinguish and contain the spread of malware. The study of disease transmission models are more centered on the quantity of traded off hosts and their dispersions, and they have been investigated widely in the software engineering group utilized a vulnerable tainted (SI) model to foresee the development of Internet worms at the early stage and as of late utilized a helpless contaminated recuperated (SIR) model to portray versatile infection engendering. One basic condition for the plague models is an extensive defenseless populace on the grounds that their guideline depends on differential conditions. More points of interest of pestilence displaying can be finding as pointed by the discoveries, which we remove from a set of watched information, ordinarily reflect parts of the contemplated objects. It is more dependable to extricate hypothetical results from fitting models with affirmation from adequate certifiable information set tests. We rehearse this guideline in this study.

II. EXISTING AND PROPOSED ALGORITHM

A. Existing System

The epidemic theory plays a leading role in malware propagation modeling. 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.

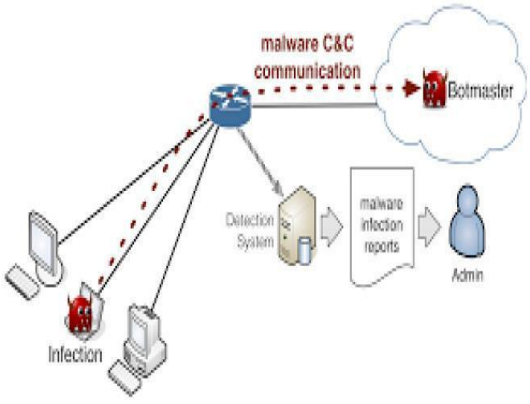

Fig.1. System Architecture of Proposed System.

B. Proposed Algorithm

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 as shown in Fig.1. Secondly, for a compromised net-work, we calculate how many hosts have been compromised since the time that the network was compromised.

III. PROBLEM STATEMENT

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 modeling. Moreover, the proposed two layer model offers us the distribution of malware in terms of the low layer networks. Future work, we will 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, e.g., the distribution of a given malware at their ISP domains, where the conditions for the two layer model may not hold.

A. Implementation of Modules

In Malware propagation in large scale networks we have the modules such as discussed below.

- Malware,
- Propagation.
- Power law

1. Malware: 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.

2. Propagation: Propagation takes place in three stages such as given below,

Early stage: An early stage of the breakout of a malware means only a small percentage of vulnerable hosts have been

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

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

3. Power Law Distribution: 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 phenomenons, 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. 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?

IV. 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. Different from Android malware, the Conficker worm is an Internet based state-of-the-art botnet. 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 1260 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 smart phones share these vulnerabilities form 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. We sort the malware subclasses according to their size (number of samples in the data set), and present them in a log format in Fig.2, the diagram is roughly a straight line. In other words, we can say that the Android malware distribution in terms of networks follows the power law.

We now examine the growth pattern of total number of compromised hosts of Android malware against time, namely, the pattern of I (t). We extract the data from the data set and present it in Table 1. We further transform the data into a graph as shown in Fig.3. It shows that the member recruitment of Android malware follows an

compromised, and the propagation follows exponential distributions.

TABLE 2. Statistics for Conficker Distribution in Terms of Ass

Number of ASes	Largest botnet	Smallest botnet
1,0048	2,825,403	1

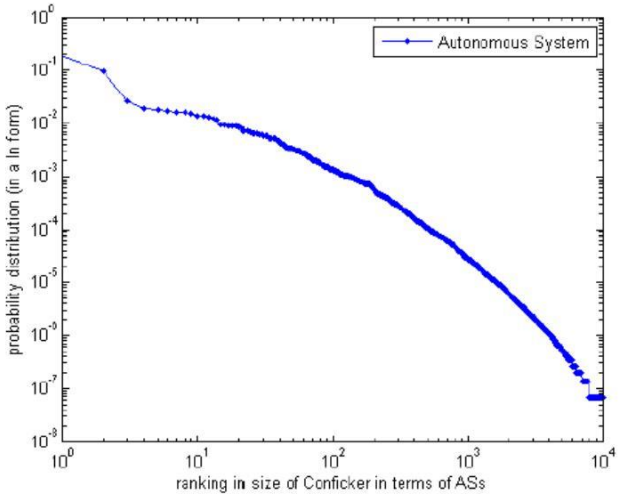

Fig.4. Power law distribution of Conficker in terms of autonomous networks.

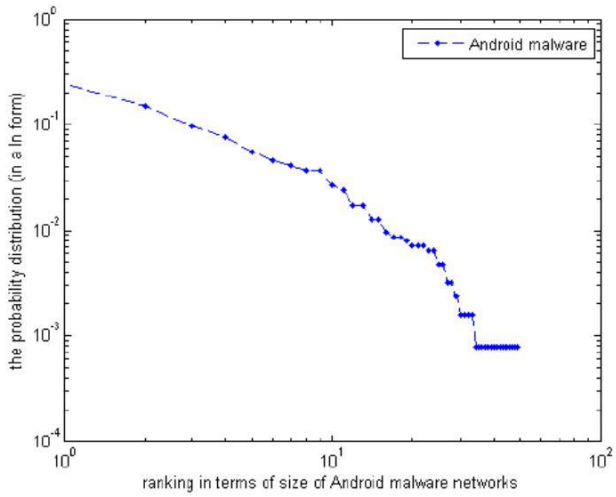

Fig.2. The probability distribution of Android malware in terms of networks.

TABLE 1. The Number of Different Android Malware Against Time (Months) in 2010-2011

Time point	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Variants	13	26	39	53	71	94	127	193	259	374	583	986	1,513	2,191	3,451

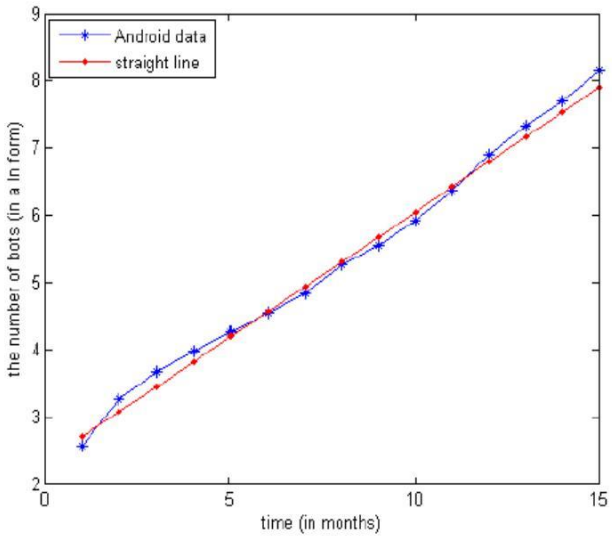

Fig.3. The growth of total compromised hosts by Android malware against time from August 2010 to October 2011.

exponential distribution nicely during the 15 months time interval. We have to note that our experiments also indicate that this data does not fit the power law (we do not show

TABLE 3. Statistics for Conficker Distribution in Terms of Domain Names at the Three Top Levels

	Number of botnets	Largest botnet	Smallest bot
top level	462	2,201,183	1
level 1	20,104	1,718,306	1
level 2	96,756	1,714,283	1

TABLE 4. The Last Six Elements of Conficker Botnet from The Top Three Domain Name Levels

	t=1	t=2	t=3	t=4	t=5	t=6
top level	9	14	18	15	22	68
level 1	543	686	924	1,534	2,972	7,898
level 2	3,461	4,085	5,234	7,451	13,002	33,522

A few key statistics from the data set are listed in Table 2. We present the data in a log format in Fig.4, which indicates that, the distribution does follow the power law. A unique feature of the power law is the scale free property. In order to examine this feature, we measure the compromised hosts in terms of domain names at three different domain levels: the top level, level 1, and level 2, respectively. Some statistics of this experiment are listed in Table 3. Once again, we present the data in a log format in Fig.5 (a), (b) and (c), respectively. The diagrams show that the main body of the three scale measures is roughly straight lines. In other words, they all fall into power law distributions. We note that the flat head in Fig.5 can be explained through a Zipf-Mandelbrot distribution. Therefore, Theorem 2 holds. In order to examine whether the tails are exponential, we take the smallest 6 data from each tail of the three levels. It is reasonable to say that they are the networks compromised at the last 6 time units, the details are listed in Table 4 (we note that t = 1 is the sixth last time point, and t = 6 is the last time point). When we present the data of Table 4 into a graph as shown in Fig.6,

we find that they fit an exponential distribution very well, especially for the level 2 and level 3 domain name cases. This experiment confirms our claim in Theorem 3.

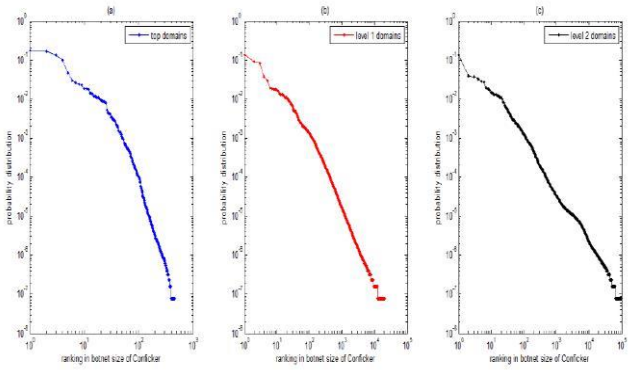

Fig.5. Power law distribution of Conficker botnet in the top three levels of domain names.

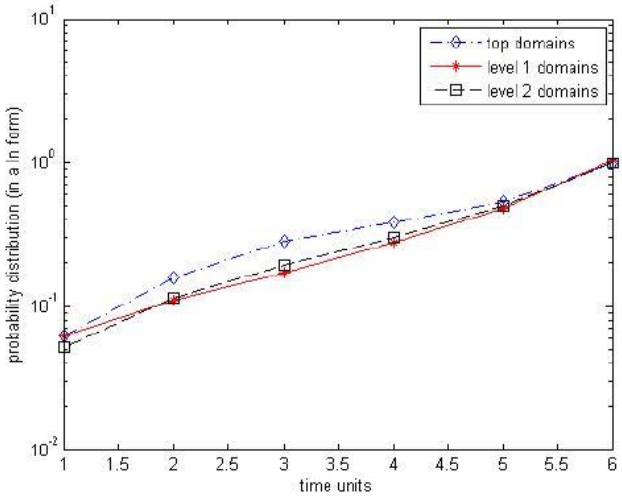

Fig.6. The three tails from the three domain name levels fit exponential distributions.

V. CONCLUSION

In this paper, we completely investigate the issue of malware appropriation everywhere scale systems. The answer for this issue is urgently wanted by digital guards as the system security group does not yet have strong answers. Not quite the same as past displaying strategies, we propose a two layer scourge show: the upper layer concentrates on systems of an expansive scale system, for instance, spaces of the Internet; the lower layer concentrates on the hosts of a given system. This two layer model enhances the exactness contrasted and the accessible single layer scourge models in malware displaying. In addition, the proposed two layer model offers us the dissemination of malware as far as the low layer systems. We perform a limited examination in light of the proposed display, and acquire three conclusions: The circulation for a given malware regarding systems takes after exponential dissemination, power law conveyance with a short exponential tail, and power law dispersion, at its initial, late, and last stage, separately. Keeping in mind the end goal to analyze our hypothetical discoveries, we have led broad analyses taking into account two certifiable huge scale malware, and the results affirm our hypothetical cases.

VI. REFERENCES

[1] Shui Yu, Senior Member, IEEE, Guofei Gu, Member, IEEE, Ahmed Barnawi, Member, IEEE, Song Guo, Senior Member, IEEE, and Ivan Stojmenovic, Fellow, IEEE, "Malware Propagation in Large-Scale Networks", IEEE 2015.

[2] 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.

[3] 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.

[4] 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.

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

Author's Profile:

J.Gomathi Dept of MCA, M.Kumarasamy College of Engineering, Karur, Tamil Nadu, India,
E-mail: gomathirithick96@gmail.com.

S. Kayathri MCA, PGDAN., Associate Professor, Dept of MCA, M.Kumarasamy College of Engineering, Karur, Tamil Nadu, India.