International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2013): 6.14 | Impact Factor (2015): 6.391

Preventing Malware Proliferation in Large Scale Network Topology

N Chandra Sekhar Reddy[1], Dr. Purna Chandra Rao[2], Dr. A Govardhan[3]

[1]MLR Institute of Technology, Hyderabad, Telangana, India

[2]Swamy Vivekananda Institute of Technology, Secunderabad, India

[3]Jawaharlal Nehru Technological Institute Hyderabad, Hyderabad, Telangana, India

Abstract: ***Malware is pernicious software in networks and spreading widely throughout in network area and cause a critical threat to network security. Till date we are not aware of malware behavior in networks. In this paper, we will find how malware propagates in networks and build a two layer epidemic model for malware propagation. In proposed model our analysis says that distribution of a given malware follows exponential distribution and power law distribution with a short exponential tail and power law distribution at its early, late and final stages and experiments also done by taking two real world malware data sets and results confirm our theoretical values.***

Keywords: Malware, Prevention, Power law, Network topology, Security

1. Introduction

Malware are pernicious software programs used by cyber attackers, by using their software they can undermine as many networked computers as they can do. An undermine computer is called as bot. All undermine bots by a malware is called as botnet. Cyber attackers constitute challenges to cyber defenders so defenders must know malware behavior such as size of botnets, distribution of bots.

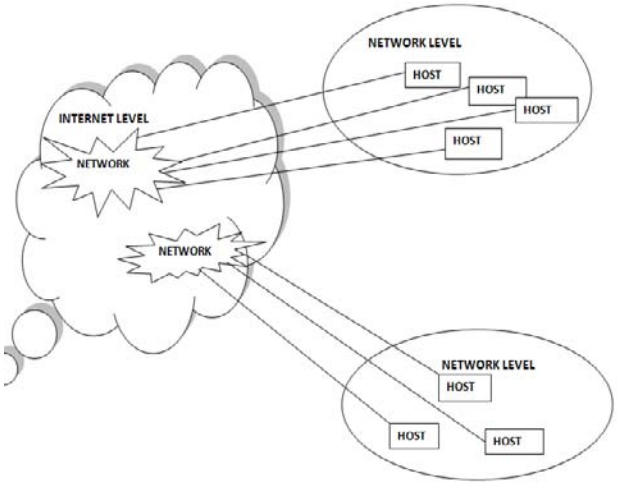

Figure 1: Internet and network level distribution

Above fig 1 indicates two levels one is internet level and other is network level were these two levels describes the malware expansion. In network level a network could be defined in many ways it could be a country network, the group of specific mobile devices and so on .In internet level every network act as an element for network level .

Malware becomes one of the most security problems on the internet and day by day it increasing and strengthen by cyber attackers. Everyone is aware about virus which is spreading in a network. When we attach new device to our computer or download anything we check for virus if it exists we use antivirus for that so it won't spread to other software in our computer. Here we have two types of nodes one is infected node and other is healthy node. If any node is infected then its spread its infection to other neighbor nodes. After the infected node is cured along with healthy node it may get attack by virus. Some models are susceptible-infected-susceptible (SIS) and susceptible-infected-removed (SIR) are used for computer virus infections.

For networks malware attacks are very important because malware is pernicious software that includes viruses, worms, and a spyware and bots .The main characteristic of malware is self expansion. The attackers mainly attack the IP address using some scanning methods and collect information about networks and forms malware. Localized scanning are used to search local networks for vulnerable hosts .Routable scanning malware are type of malware they select target only in routable address space.

In this paper it describes how attackers are attack the information of users so that they can take control over on that system here we have the community where they can control the system and see that no one attack the systems if they find any system attack they recover the system using model.

2. Background Work

To date, we are not having any strong believed of malware size and distribution botnets. Researchers have appointed different methods to measure the size of botnets, such as botnet infiltration [1], External information [2], DNS redirection [3]. These attempts indicate that the size of botnets varies from millions to a few thousands. Researchers seriously desire effective models and explanations for the chaos[12]. Dagon et al [3] revealed that number of available bots have obvious impact by time zone. In survey paper [8] we can find more details about mobile malware. Mieghem et al [4] indicates that malware spreading has more impact by network topology through their rigorous mathematical analysis. Finding about malware propagation in network topology by Chen and ji [9]: the distribution is non uniform.

This indicates in this field he has done research in its early stage.

The current models for malware divided in to two types the epidemiology model and control theoretic model. The control system theory based models try to detect the spread of malware [10].The epidemiology model are more concentrate on number of undermine hosts and their distributions said by S.H.Sellke, N.B.Shroff and S.Bagchi. The epidemic theory plays a major role in malware propagation. The epidemic models have one critical condition that is a large vulnerable population because their principle is based on different equations. We can find more details about epidemic model by D.J.Daley and J.Gani. Zhou et al by using Susceptible-infected (SI) model at early stage he predict growth of internet worms .A Susceptible-infected-recovered model (SIR) used by Geo and Liu to describe mobile virus propagation. The Distribution of malware in large scale networks has a sufficient volume of data to meet the requirements of the SI model.

Apart from epidemic model, we divided our model in to two layers. First, for a given time we calculate how many networks have been compromised based on SI model. Second, for undermine network we calculate how many host have been undermine from the time that network have been undermine. From this analysis we find that the distribution of given malware follows exponential distribution at early stage, and accept power law distribution with short exponential tail at its late stage. The two layer epidemic model better in large scale networks compared with single layer epidemic model. We determine our theoretical values through two real word data sets: The Android based malware Y.Zhou and X.Jiang and conficker [11] and confirmed that theoretical values are same as experimental values.

A program which is written by malware programmer, called as bot or agent. Combination of bots is called as botnet. Using Virus techniques install bots at undermine computers on the internet. Botnets were controlled by owner to perform illegal tasks. To communicate with the bots and collect data from that we have command and control (c & c) server(s). The URL of the C & C was regularly changed by botmaster to divert from legal forces. We get more information about this from [1]. Now- a- days many mobile malware increased along with smart phones. The Symbian OS for mobile devices was first developed by Cabir [5]. Using Bluetooth it was the first malware propagating. Malware against Apple iphones was developed by Ikee [6]. Malware against Windows CE OS was developed by Brador [7]. The history of mobile malware and surveyed their propagation models by Peng et al [8].

Torpig botnet URL was register by stone Gross et al [1] before botmaster to hijack the C&C server for ten days in that duration they will collect 70GB data from the bots, and they reported that Torpig botnet was 182,800 were median and average size of Torpig's live population was 49272 and 48532 and they indicated that live populations fluctuates as user switch between being online and offline. Some issues on this were given by Dagon et al in [3].

Botnets are major source for cyber criminals to carry out their prohibited work. Such as sending spam mails by A.Ramachandran in [1], taking personal data such as mail accounts or bank credentials by T.Holz and S.Saroiu in [1]. We can get more details about bonnet analysis and size from [1]. The size of botnet were became debate for research community. The hidden botnet connections were finded and exposed in [2] and discussed their hidden relationships among botnets. From [2] we came to know that, they first create a botnet dimensional structural feature vector. For unique identity we have the following features: DNS name or IP address of IRC server, name of the IRC network name, IRC Channel name, server version and Botmaster IDS (We can get IDS from IRC trace.

The observations of previous botnets help us to predict the future botnets. Botnets are widespread so we need a technique to rank them .Existing models predict the total botnet population but it will takes lengthy period of time.[12] Virus will take less time to spread infections so we need a model to predict population growth in less time .The computers in every time zone form as a group . The computers in each time zone whether they are infected or still vulnerable have the same diurnal dynamics. The number of infected hosts in a region may vary according to time. We can know more details about time zone in [3].

3. System Architecture

In malware we can see mainly following interactions that is Admin, user and Database that interaction were shown in below fig 4. The Admin Server will perform the operations like analyzing documents and it will check the content of the document whether that document contain malware are not. If the documents contain malware then Admin will scanned that documents to find the malicious users who will increase malware in social networks and admin will keep them in block list.

Data provider will play a key role in this system design because it will browse the required file and uploads in the social network; after uploading it will share with their friends. The data provider will perform the following operations like Add documents and view documents. Any user wants to add new document they must enter document title and document name.

After submission it will store in the database .If data provider wants to view documents details like document name, document title, and document content and related images.
User is the main key role in the system design because user will perform the operations like view messages, send messages, view or search users and send friend request and friends can share data among them. User can do these types of operations after login with their user id and password .User will get user id and password after successful registration .Registration contain user details.

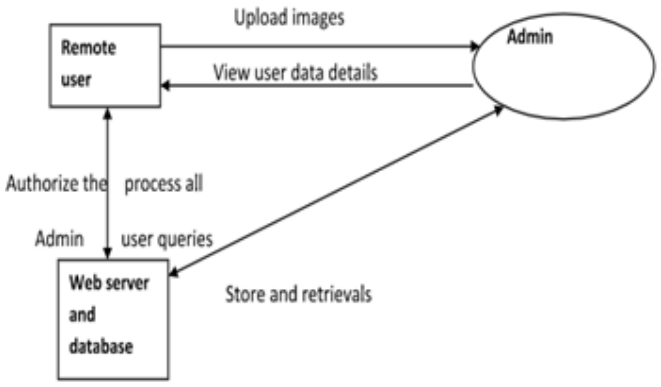

Figure 2: Interaction between user and admin

4. Proposed Methodology

Early stage : - At early stage , the breakout of malware means only the less percentage of vulnerable hosts have been undermine and follows exponential distributions .
Final stage: - At Final stage, malware means all vulnerable hosts have been undermine.
Late stage: - At late stage, the time interval between both previous stages were finded.

We examine our theoretical analysis through two large-scale malware: Android malware and conficker. In recent Android malware is a fast developing leading smart phone based malware by Zhou and jiang. Conficker is the most recent worm; it is the internet based state-of-the-art botnet [11]. As reported in [11], it has infected about 7 million to 15 million hosts and till now victims increasing .By studying conficker, we aspire to understand the current and new trends in malware propagation , it will help to predict in future malware trends . We came to know that conficker has very different victim distribution patterns compared to previous generation botnets. When we faced with new malware threats such as conficker at that time reputation-based blacklisting approach can perform but by cross checking several DNS blacklists and IP/AS reputation data from DShield , FIRE and STUDY in [11] shows that a blacklist-based approach can detect most bots and reputation-based were poorly detected . Reputation based techniques for future malware defense shown that neighborhood watch is an effective approach in conficker case. From this we came to know that security alert play a better solution for future malware defense. The important feature of conficker is, it will update itself. It automatically generates new domain names D.Watson in [11] and it wills connects to those domain names and update its version automatically to avoid detection.

Theorem: 1
At the early stage of malware propagation in the internet, the malware propagation follows exponential distributions.

Proof:
At a time point of the early stage ($0 \le t < T_e$) of a malware breakout the number of compromised networks as

$$C(t) = C(0)e^{\beta N t}$$

C (t) follows exponential distribution. For any of the compromised networks, it has progressed t' ($0 < t' \le t < T_e$) time units and its size is

$$S(k_i, t') = C_i(0)e^{\beta M_i t'}$$

We find that, at the early stage all the size of compromised networks follows an exponential distribution.

At final stage we can get concrete conclusion of the propagation of malware.

Theorem: 2
At the final stage (t=∞) of malware propagation in the internet the malware propagation follows power law distribution.

Proof:
All the vulnerable hosts have been compromised namely

$$S(k_i, \infty) = M_i, i = 1,2,\ldots,N.$$

Based on the above discussion, We know M_i(i=1,2,…,N) follows the power law distribution.
Now we move to late stage of malware propagation.

Theorem: 3
At the late stage ($T_e \le t < \infty$) of a malware breakout in the internet, the malware distribution contains two parts namely power law and short exponential tail.

Proof:
Assume that a malware propagation has progressed for t (t>>T_e) time units. Let t'=t-T_e . The compromised C (t) hosts by time point t' were separated then we have two groups of compromised hosts.

According to theorem 2, the compromised hosts before t' follows the power law as t'>>T_e and all the compromised hosts after t' are still in their early stage .Hence these compromised networks follows Exponential distribution.
The network compromised after time point t' are at the tail of the distribution.

For a given network K_i for $t_1 > t_2$ we have,

$$S(k_i, t_1) \ge S(K_i, t_2)$$

Due to the fact that t'>>
T_e , *based on this fact statement is given as* t he length of exponential tail is much shorter than the length of main body.

Experiments
For Experiment purpose in this we use two malware data sets. Shin et al [11] from all over the world he collected about 25 million conficker data sets at different levels. The large data set about Android based malware were collected by Zhou and jiang . Stochastic branching process model were proposed by Sellke et al for propagation of internet worms characteristics it will mainly focus on the number of undermine computers against the worms scan number. Ganesh et al fined that speed of epidemics were affected by network topology. To analyze the spread of virus in networks N-intertwined Markov chain model were applied by Mieghem e al. [4]. In below fig 3, we came to know how time increasing according to recruited members

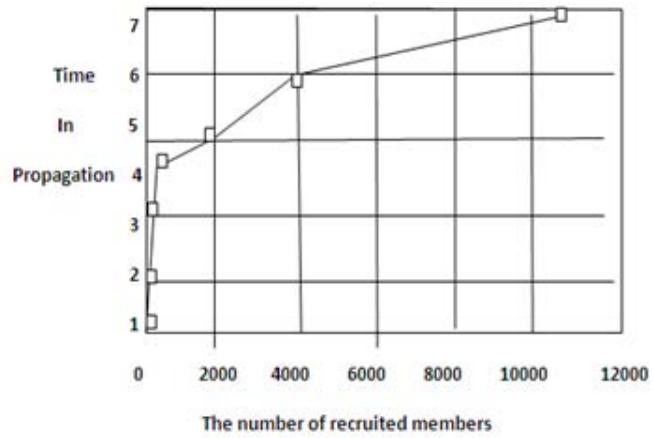

Figure 3: Time propagation for recruited members

5. Experimental Results

We have created a user log in and Admin log in to upload files into database. Where user login and upload his file as shown in fig:4.

Figure 4: Upload files

Figure 5: Admin login

Fig 5: depicts the login page for admin where admin can login and view the uploaded files and provide security. Admin can view the friend requests and responses as shown in fig: 6

Figure 6: View user friend request and response

User Name	File Name	Picture	Title	Uses	Description FileName	Description	MAC	SK	Rank	Date
swathi	img2.jpg		cloud	cloud	Lalbag.txt	Lalbagh park or Lalbagh Botanical Gardens, meaning The Red Garden in English, is a well	b5f811e9e8d14415efbee77a58c9ccb24e1b1a8	[B@30c06794	0	27/09/2016 15:25:46
swathi	gal1.jpg		user file	swathi	Bannergatte.txt	Bannerghatta National Park, near Bangalore, Karnataka, was founded	6b1a9d9b146aaa2213cbfd6ef079fc8781df75fa	[B@206564e9	0	27/09/2016 15:43:54

Figure 7: View all documents

Admin can view all the documents in database as shown in fig: 7

Figure 8: Block user

The admin can block any user who is trying to do any harmful activity as shown in fig: 8. Hence security can be enhanced.

6. Conclusion

In this paper we completely known the problem of malware distribution at large scale networks and solution to this problem is given by cyber defender as the network security community. Comparing with single layer epidemic model, this two layer epidemic model improves accuracy. In this two layer, upper layer concentrate on the networks of a large scale and lower layer concentrate on the hosts of a given network .After performing the restricted analysis based on the proposed model we obtain three conclusions :The given malware in networks follow exponential distribution ,power law distribution at its early, late and final stage . After performing experiments by taking two real world large scale malware their results conforms theoretical values.

7. Future Work

In further discussion we have many directions that could be further explored. Some of them are listed below:

- 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 reality at the same network multiple malware may coexist but the fact that different malware focus on different vulnerabilities, and distribution of different malware may not be same.
- We can find more details about the length of exponential tail of a power law distribution at late stage. Defenders can concentrate more on networks.
- In this paper we focused only on one malware, In future we are interested to find distribution of multiple malware on large scale networks.

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 Proc. ACM Conf. Comput. Commun. Security, 2009, pp. 635–647.

[2] 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 Proc. 1st Conf. 1st Workshop Hot Topics Understanding Botnets, 2007, p. 5.

[3] D. Dagon, C. Zou, and W. Lee, "Modeling botnet propagation using time zones," in Proc. 13th Netw. Distrib. Syst. Security Symp., 2006.

[4] P. V. Mieghem, J. Omic, and R. Kooij, "Virus spread in networks," IEEE/ACM Trans. Netw., vol. 17, no. 1, pp. 1–14, Feb. 2009.

[5] Cabir. (2014). [Online]. Available: http://www.f-secure.com/en/ web/labs_global/2004-threat-summary.

[6] Ikee (2014) Available: http://www.f-secure.com/vdescs/worm_iphoneos_ikee_b.shtml

[7] Brador. (2014). [Online]. Available: http://www.f-secure.com/vdescs/brador.shtml

[8] S. Peng, S. Yu, and A. Yang, "Smartphone malware and its propagation modeling: A survey," IEEE Commun. Surveys Tuts., vol. 16, no. 2, pp. 925–941, 2014.

[9] Z. Chen and C. Ji, "An information-theoretic view of network aware malware attacks," IEEE Trans. Inf. Forensics Security, vol. 4, no. 3, pp. 530–541, Sep. 2009.

[10] A. M. Jeffrey, X. Xia, and I. K. Craig, "When to initiate HIV therapy: A control theoretic approach," IEEE Trans. Biomed. Eng., vol. 50, no. 11, pp. 1213–1220, Nov. 2003.

[11] S. Shin, G. Gu , A . L. N. Reddy, and C. P. Lee,"A large scale empirical study of conficker," IEEE Trans. Inf.Forensics Security, vol. 7,no. 2,pp.676-690,Apr.2012.

[12] Shui Yu,senior member ,IEEE,Guofei Gu,Member,IEEE,SongGuo ,senior member,IEEE "Malware propagation in large scale networks"IEEE Transactions on knowledge and data engineering ,vol .27 ,no .1,january 2015.