Restraining Hybrid Mobile Virus Propagation

S.Rajkumar[1], P. E. Rubini[2]
[1]ME, [2]Asst. Professor
Computer Science Engineering, SRM University, Ramapuram Campus

Abstract

Mobile Virus propagation has been increased with rapid growth of smart phones. Proportionately it becomes increasingly difficult to secure them against attack in the form of virus, which may disturb the operation as well as the function of the mobile phones. It can cause system failure, wasting memory space, extra power and data leakage. Hence forbidding mobile virus propagation is very important one. SEIR model was proposed to restrain hybrid spread with less cost. N-gram analyzer to filter the incoming virus message in the mobile and the virus message is stored in separate folder in mobile and this will reduce the traffic of mobile network compared to existing method.

Keywords

SEIR, hybrid spread, n-gram analyser,

1. Introduction

Nowadays mobile phones are used in many activities like Net banking, online shopping, Education, Business, Entertainment, Media etc., because of this high-end facilities, mobile phones are very prominent than computers. The outcome of the mobile phone usage is to escalate the dishonest people who would like to take an advantage of these actions for unlawful gains. Malware is a very big danger in the present technology driven world. Mobile Malware is an acronym of "malicious software" - particularly built to target mobile devices such as smart phones, tablets to harm the devices. Malware has the capability to contaminate other system files, executable files and corrupts the data.The various attack channels used by smart phone virus, such as Bluetooth, SMS/MMS and so on. The impact of the multi-channels propagation of smart phone virus is even more serious than the single mode, so it is necessary to study the multi-channels spread of smart phone virus.

In previous works some model has been proposed for dynamic propagation of virus. Valid propagation models can be used as test beds to estimate the scale of a virus outbreak before it occurs in reality [6] and to evaluate new and/or improved counter measures for restraining virus propagation [7]. Recently, there exist some models to characterize and predict the infection dynamics of mobile viruses [8], [9], [10]. However most of the works considered only human behaviour and for human mobility they used levy flight and random walks. Moving probability and revisit probability to old place is not considered.

SEIR model was proposed to restrain mobile virus propagation. Viruses are triggered as a result of human behaviours, rather than contact probabilities in homogeneous model [9]. The two types of human behaviours are operational behaviour and mobile behaviour. The human may open the virus message is considered as operational behaviour. The movement of user is considered as mobile behaviour. Different from existing work, SEIR model focused on hybrid spread of virus and differentiate the infected message and not infected message to the user.

Many methods have been proposed recently to restrain mobile virus propagation. However, these methods will protect infected mobile from sending infected message based on system call sequence and API [4], [5], [15]. Existing system will not able to identify the new virus, because of the limitation of the antivirus knowledge. In order to update the detection database timely in user mobile the patch is disseminated to all smart phones by service provider or security companies. It is not possible to send patch or notification to all mobile because of the limitation of band width and time. Some strategies have been used to disseminate patch but it is not suitable for large scale network.

Two ways propagation is used in existing model to restrain mobile virus. The two way propagations are Bluetooth based and SMS/MMS based. The Bluetooth based is short range propagation and SMS/MMS based is wide range propagation, it will send infected message to all the contacts in the particular mobile. In order to restrain the propagation of virus message patch is disseminated to all mobile but this method

will increase traffic and the cost.Thus, a two way securities for restraining hybrid mobile virus was proposed. The two way securities are SEIR model to restrain hybrid virus propagation and n-gram analyzer to differentiate infected and not infected messages so probability of opening the infected message will be reduced. The autonomy oriented computing (AOC) is used for pre immunization. The mobile are differentiated in three ways infected mobile, most likely to infect and secured mobile. The patch is disseminated to most likely to infect mobile first and then to others.

2. Related Works

2.1. Characteristics of Bluetooth network and SMS/MMS network

Each phone can communicate with others through Bluetooth, which form Bluetooth network automatically. In the network each phone is defined as a node. The effect of node connectivity to Smartphone virus cannot be ignored, the influence factors of node connectivity conclude: Bluetooth signal coverage radius, density, and moving speed and so on, which can represent these factors through defining node average degree [5]. Stronger node connectivity, faster the virus propagation speeds.

The connectivity of each phone through SMS/MMS forms SMS/MMS network, In the SMS/MMS network there are hundreds of millions of mobile phone users, it also has a strong aggregation characteristics [8]. The greater of users gather density, which means more frequent contacting between users, the better spread of Smartphone viruses.

2.2. Virus propagation through BT and SMS

According to the communication channels of mobile viruses, the viruses fall into two categories namely: BT-based viruses (e.g., Cabir, Lasco) and SMS-based viruses (e.g., TXSBBSpy, Zombie, and Commwarrior). A BT-based virus is a local-contact driven virus since it infects other phones only through Bluetooth and Wi-Fi devices within a given radio range. Similar to contact based diseases as in humans (e.g., SARS and H1N1), the propagation of a BT-based virus follows a spatially localized spreading pattern. Epidemic modelling is one of the most common approaches for studying such virus propagation. It assumes that individuals are homogeneous in a host community, each having an equal likelihood contact with others. Also, epidemic modelling is applied on some studies to analyze the propagation dynamics of a BT-based virus. SMS-based viruses can send copies of themselves to all phones that are recorded in address books, by means of photos forwarding, videos, and short messages, etc. The propagation of SMS-based malwares follows a long-range spreading pattern that is similar to the spreading of viruses in computer, especially like worm propagation in e-mail networks thus, the operational behaviour of users is important in SMS-based virus propagation. Users with awareness about the viruses risk will not likely be infected even if they receive attachment. In order to study SMS/MMS-based virus propagation, certain operational patterns are considered, such as if the users open a virus attachment or not.

On the other hand, SMS-based viruses can send copies of themselves to all phones that are recorded in address books, by means of forwarding photos, videos, and short messages, etc. The propagation of SMS/MMS-based viruses in mobile networks follows a long-range spreading pattern that is similar to the spreading of computer viruses, especially worm propagation in e-mail networks [7], [13].

Related research on human mobility and operational behaviour are incorporated into SEIR model in order to provide a computational model for characterizing and simulating the propagation dynamics of mobile viruses. The traits of mobility patterns described in SEIR model are consistent with statistical results from the real-world traces, i.e., local bounded mobility areas; power-law travelling distances, and inters contact times.

2.3. Defence Strategies against Mobile Viruses

Some countermeasures such as anomaly detection technologies have been proposed to protect users' private information from being revealed to different users. Like, Bose et al. discriminated some of the malicious behaviours from normal operations by training a classifier based on the method of support vector machines. Cheng et al. have provided an approach to detecting both single-device and system-wide abnormal behaviours by collecting and sending communization data to remote servers in order to reduce the detection burden of phones.

Although these abnormal detection technologies can help directly protect phones from being affected by certain viruses, it is not easy to detect new viruses because the monitoring technologies must first be trained to recognize normal and abnormal operational behaviours. If any new virus produces some patterns (e.g., a series of system calls), these monitoring technologies cannot detect such virus. Hence, challenging to detect a worm outbreak at the early stage unless both users and security companies frequently update their detection classifiers. Different from wired networks (e.g., computer networks), it is almost impossible to send patches to all phones simultaneously and timely. Thus, new strategies was needed to efficiently disseminate security notifications or

patches to as many phones as possible with a relatively lower communication cost before a new virus spreads to a large population. In order to reduce communication redundancy, strategies that send patches based on Bluetooth is utilized. After which they send security signatures to all communities based on the local detection. However, the method cannot ensure that users acquire patches in time. The performance of an AOC-based pre-immunization strategy is examined and that selects some highly-connected phones and prevents a virus from turning into an epidemic. Furthermore, AOC-based dissemination strategy is designed that distributes security notifications or patches to smart phones with a low communication redundancy, in order to restrain virus propagation before it causes further infections.

2.4. Modelling and restraining mobile virus propagation

A two-layer network model for characterizing BT-based and SMS/MMS-based viruses which propagate through Bluetooth and Short/Multimedia Message Services. BT based propagation will propagate only to short range. The SMS/MMS based viruses will spread to all the contacts in the mobile so it will spread to wide range. The patch is send by two methods through SMS/MMS and Bluetooth but it is not sure that patch will send in time. Operational behaviour such as probability of user open the infected message and mobility behaviour represent the mobility of user and AOC is used for pre-immunization, method doesn't consider about hybrid virus propagation.

3 Proposed System

In proposed system, SEIR model is used which will restrain the mobile virus propagation. AOC is used to detect the virus automatically and fix it. The n-gram analyzer is used to differentiate virus message and the not infected message. The virus message is stored in bin. This method will restrain the hybrid virus propagation. The mobile is clustered into three infected, secured and most likely to infect. The patch is send to the most likely to infected mobile first and then from that mobile to others through Bluetooth and SMS.

3.1. Propagation of Patch

The patch is disseminated to the mobile through Bluetooth and from service provider. The service provider will send patch to the suspectible mobile first and then from that mobile to other mobiles through Bluetooth/WI-FI. By using this method cost will be reduced shown in Fig.3.1

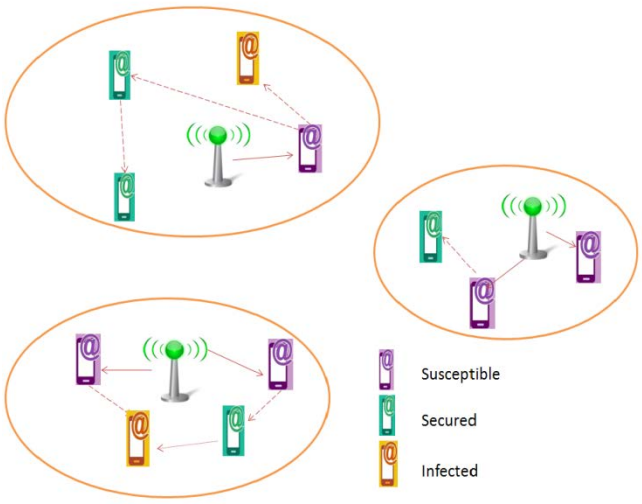

Fig. 3.1.Propagation of patch

3.2. SEIR Model

In SEIR model, nodes have been divided into four status: Susceptible (S), Exposed (E), Infected (I), Removed/Recovered (R). The Fig. 3.2 shows the working of SEIR model.SEIR model based on the proceeding assumptions are listed as follows:

$$\frac{ds(t)}{dt} = -\propto S(t)I(t) + \mu I(t)$$

$$\frac{dE(t)}{dt} = -\beta E(t) + \alpha S(t)I(t)$$

$$\frac{dI(t)}{dt} = \beta E(t) - \gamma I(t) - \mu I(t)$$

$$\frac{dR(t)}{dt} = \gamma I(t)$$

Where S(t) is used to represent the number of the nodes not yet infected with the disease at time t, or they are susceptible to the disease; E(t) denotes the number of nodes who have been infected with disease, but not have been capable of propagating the disease to those in the susceptible category, R(t) is used to denote the number of the nodes who have been infected and then recovered from the disease. β is the infection rate, α denotes the probability when the exposed change to the infection status, γ is the rate when susceptible become the recovered nodes, μ represent the rate of infection changing to susceptible status again.

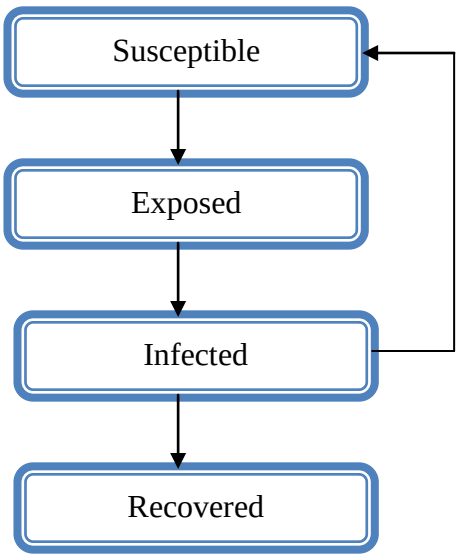

Fig.3.2. Transition between nodes

3.3. N-GRAM Analysis

N-gram model is a type of probabilistic language model for predicting the next item in such a sequence in the form of a (n-1) order Markov model. N-gram models are now widely used in probability, communication theory, computational linguistics and data compression. It has the capability to capture the inherent features of the given input data. It is used to extract the most frequent N-gram signatures in the given database. When a new code is analyzed, it can be classified as malicious message and un infected message based on the category it matches the most. An advantages of N-gram models are relatively simplicity and ability to scale up by simply increasing n-model can be used to store more context with well understand space time trade off, enabling small experiments to scale up very efficiently.

Using this n-gram analyzer the message are differentiated and store the infected message in separate folder. So that user won't open the infected message. The document frequency is a kind of global measure the class wise document frequency is a local measure with respect to a class. The main advantage of class wise document frequency is that each class can analyzed independently. Moreover, the number of distinct *n*-grams in a program is far smaller than that for the entire training set and hence it saves memory requirements and deal with only *n*-grams of one class at a time. A novel concept of relevant n-grams for a class D of executable programs. For each executable program, the set of all n-grams can be find and let Ng^t be the set of all n-grams for a program t $\in$ D.

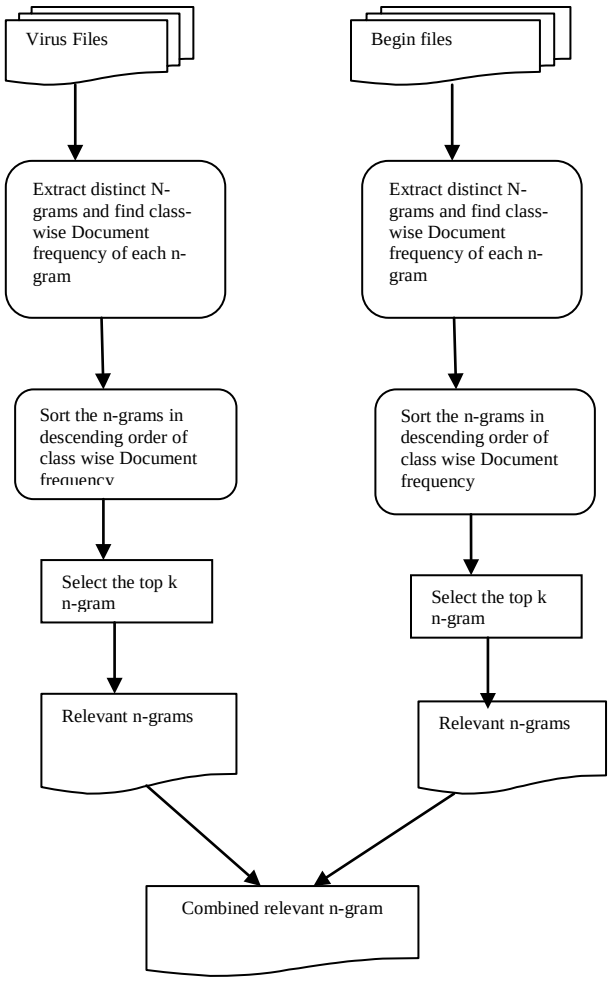

Fig.3.3 N-gram flow chart

Once the set of relevant n-gram are selected, the vector-space model of information retrieval is used to represent the set of program D in term of $Ng_k(D)$. A program is represented as a vector of terms $t_1, t_2 \ldots t_m$ where $t_f(1 \leq i \leq m)$ is a binary (0-1) value denoting the single or multiple occurrences of i^{th} n-gram $Ng_k(D)$ in the program. A program is viewed as a vector each unique relevant n-gram corresponds to a dimension. The value 1 represents the occurrences of the *n*-gram in the program and its absence is represented by 0. Training set consists of a set of labelled vectors— the vector representation of the set of programs together with the respective class label. Fig.3.3 shows the flow of files.

3.4. Autonomy-Oriented Computing

AOC has three goals [10], the first goal is to reproduce life-like behavior in computation. With complete knowledge of the fundamental mechanism, simplified life-like behavior can be used as model for a general-purpose problem solving technique. Replication of behavior is not the end, but rather

the means, of these computational algorithms; the second goal is to understand the essential mechanism of a real-world complex system by hypothesizing and frequent experimentation. The conclude product of these simulations is a progress understanding of or explanations to the real working mechanism of the modeled system; the third goal affairs the rise of a problem solver in the absence of human intervention. To build an AOC-based model, the following is a list of common steps:

- Observe macroscopic behaviors of a natural system;
- Design entities with desired synthetic behaviors as well as an environment where entities reside;
- Observe macroscopic behaviors of the artificial system;
- Validate the behaviors of the artificial system against the natural counterpart;
- modify (ii) in view of (iv);
- Repeat (iii)-(v) until satisfactory;
- Find out a model/origin of (i) in terms of (ii) or apply.

From the above steps, note that an AOC system mainly contains a population of autonomous entities and the rest of the system is referred to as the environment.

4. Conclusion

SEIR model was proposed and introduced relevant *n*-grams in the context of using *n*-grams for computer virus detection. The SEIR model is used to restrain the hybrid mobile virus. The *n*-gram approach lacks semantic awareness and it is very difficult to analyse the relevant *n*-grams results that obtained. The future work will be to develop a semantic aware method and also include different kinds of malicious and benign executables in our training data and also exploring the use of variable- length *n*-grams.

References

[1] Chao Gao and Jiming Liu, "Modelling and Restraining Mobile virus propagation", "Transactions on Mobile computing", vol. 12 No. 3, March 2013.

[2] D.-H. Shi, B. Lin, H.-S. Chiang, and M.-H. Shih, "Security Aspects of Mobile Phone Virus: A Critical Survey," Industrial Management and Data System, vol. 108, no. 4, pp. 478-494, 2008.

[3] H. Kim, J. Smith, and K.G. Shin, "Detecting Energy-Greedy Anomalies and Mobile Malware Variants," Proc. Sixth Int'l Conf. Mobile Systems, Applications, and Services (MobiSys 08), pp. 239-252, 2008.

[4] L. Xie, H. Song, T. Jaeger, and S. Zhu, "A Systematic Approach for Cell-Phone Worm Containment," Proc. 17th Int'l World Wide Web Conf. (WWW '08), pp. 1083-1084, 2008.

[5] N. Xu, F. Zhang, Y. Luo, W. Jia, D. Xuan, and J. Teng, "Stealthy Video Capture: A New Video-Based Spyware in 3G Smartphones," Proc. Second ACM Conf. Wireless Network Security (WiSec'09), pp. 69-78, 2009.

[6] G. Yan and S. Eidenbenz, "Modeling Propagation Dynamics of Bluetooth Worms (Extended Version)," IEEE Trans. Mobile Computing, vol. 8, no. 3, pp. 353-368, Mar. 2009.

[7] C. Gao, J. Liu, and N. Zhong, "Network Immunization and Virus Propagation in Email Networks: Experimental Evaluation and Analysis," Knowledge and Information Systems, vol. 27, no. 2, pp. 253-279, 2011.

[8] P. Wang, M.C. Gonzalez, C.A. Hidalgo, and A.-L. Barabasi, "Understanding the Spreading Patterns of Mobile Phone Viruses," Science, vol. 324, no. 5930, pp. 1071-1076, 2009.

[9] S. Cheng, W.C. Ao, P. Chen, and K. Chen, "On Modeling Malware Propagation in Generalized Social Networks," IEEE Comm. Letters, vol. 15, no. 1, pp. 25-27, Jan. 2011.

[10] P. De, Y. Liu, and S.K. Das, "An Epidemic Theoretic Framework for Vulnerability Analysis of Broadcast Protocols in Wireless Sensor Networks," IEEE Trans. Mobile Computing, vol. 8, no. 3, pp. 413-425, Mar. 2009.

[11] P. De, Y. Liu, and S.K. Das, "Deployment Aware Modeling of Node Compromise Spread in Wireless Sensor Networks Using Epidemic Theory," ACM Trans. Sensor Networks, vol. 5, no. 3, pp. 1-33, 2009.

[12] M.C. Gonzalez, C.A. Hidalgo, and A.L. Barabasi, "Understanding Individual Human Mobility Patterns," Nature, vol. 453, no. 7196, pp. 779-782, 2008.

[13] C.C. Zou, D. Towsley, and W. Gong, "Modeling and Simulation Study of the Propagation and Defense of Internet E-Mail Worms," IEEE Trans. Dependable and Secure Computing, vol. 4, no. 2, pp. 105-118, Apr.-June 2007.

[14] L. Xie, X. Zhang, A. Chaugule, T. Jaeger, and S. Zhu, "Designing System-Level Defenses against Cellphone Malware," Proc. IEEE 28th Int'l Symp. Reliable Distributed Systems (SRDS'09), pp. 83-90, 2009.

[15] A. Bose, X. Hu, K.G. Shin, and T. Park, "Behavioral Detection of Malware on Mobile Handsets," Proc. Sixth Int'l Conf. Mobile Systems, Applications, and Services (MobiSys '08), pp. 225-238, 2008.

[16] F. Li, Y. Yang, and J. Wu, "CPMC: An Efficient Proximity Malware Coping Scheme in Smartphone-Based Mobile Networks," Proc. IEEE INFOCOM, pp. 2811-2819, 2010.

[17] G. Zyba, G.M. Voelker, M. Liljenstam, A. Mehes, and P. Johansson, "Defending Mobile Phones from Proximity Malware," Proc. IEEE INFOCOM, pp. 1503-1511, 2009.

[18] P. Wang and M.C. Gonzalez, "Understanding Spatial Connectivity of Individuals with Non Uniform Population Density," Philosophical Trans. Royal Soc. A, vol. 367, no. 1901, pp. 3321-3329, 2009.

[19] A. Mei and J. Stefa, "SWIM: A Simple Model to Generate Small Mobile Worlds," Proc. IEEE INFOCOM, pp. 2106-2113, 2010.

[20] K. Lee, S. Hong, S.J. Kim, I. Rhee, and S. Chong, "SLAW: A Mobility Model for Human Walks," Proc. IEEE INFOCOM, pp. 855-863, 2009.