ISSN:0975-9646

Reinfection and Self-Start Analysis on the Propagation Dynamics of Modern Email Malware

Preetha. S.M

Department of Computer Science, CUSAT
Sarabhai Institute of science and Technology, Vellanad,Trivandrum, Kerala

Abstract— **In recent years, email is the basic service for person to person communication, and email facilitates by its high speed, and process ability. The email malware exhibits two new propagation features; reinfection and self-start. Reinfection is the process by which an infected user sends out malware copies, whenever the infected user opens the malicious hyperlink or attachment. Self-Start is the process by which the infected user spreads the malware copies, whenever certain events are triggered. To solve this problem, derive a new analytical model by introducing a concept of virtual nodes. The malware detector serves as an empirical means of evaluating malware detection techniques detection capabilities. The new analytical model can efficiently predict the reinfection detection and effectively overcome the associated computational challenges.**

Index Terms*— *Email Malware, Propagation, Malware Detection.

I. Introduction

Among Internet users, Email is considered as the most important application on the Internet. In Sweden the number of users connected to the Internet has doubled each year for several years. This growth involves new groups of users in the Email community and this makes Email usage and its impact on workplaces important to study. An extensive description of electronic mail is given in Palme [1], where the technical, legal, and economical factors are described and analysed. Email is a computer-based communication system where messages can be written by a sender on a computer. These messages are then transmitted via computers to the addressee's mail server where they can be opened and read by the receiver.

Network security is an important task of network management. One threat to network security is malware (malicious software) propagation [2]. Malware is malicious software in short, which is designed especially for either damaging or disrupting a computer system. Email worm [3], is defined as a piece of malicious code that spreads through Email by including a copy of itself in the Email attachment— an Email user will be infected if he or she opens the worm Email attachment. If the Email user opens the attachment, the worm program will infect the user's computer and send itself as an attachment to all Email addresses that can be found in the user's computer. An Email worm spreads on a logical network defined by Email address relationship; it's difficult to mathematically analyze Email worm propagation.

Protecting a computer system from malicious attacks is a key challenge to network security and management. One such attack is due to malware propagation [2]. An Email user is called infected once the user opens a worm Email attachment; upon opening a worm attachment, an infected user immediately sends out a worm Email to all neighbors' [3].

Important fact of an Email network [9] (in terms of Email worm propagation) is that once a computer contains the address of an Email list, from an Email worm's point of view, this computer has virtually all the addresses associated with the Email list. Therefore, even though a user's computer may only contain tens of Email addresses, the degree of the user in the Email network [9] might be as large as several thousand if one of the Email addresses is a popular Email list.

There are many ways to attack Emails, which affects the sending Emails (Email backscatter) i.e. spam Emails using viruses or worms. For that, so need to inform the sender about the real reasons for not receiving Email from the other side. The attackers intercept the Email, and delete the sender's address, therefore the Email gets spammed and the receiving process fails, thus the sender receives a failing note message and he/she cannot determine the real reason of failure. The Email spam propagation can be analyzed by many factors such as the period of time between sending the Email and sending back the failing report for the sender, another factor is the returned message which does not contain a real failing reason [3]. Creating reliable models of virus and worm propagation is beneficial for many reasons. It allows researchers to better understand the threat posed by new attack vector and new propagation techniques. For instance, the use of conceptual models of worm propagation allowed researchers to predict the behavior of future malware, and later to verify that their predictions were substantially correct [7][8].

Malware is referred to by numerous names. Examples include malicious software, Malicious code (MC) and malcode. Numerous definitions have been offered to describe malware. A malware detector is the implementation of some malware detection technique(s) [4]. The malware detector attempts to help protect the system by detecting malicious behaviour. The malware detector may or may not reside on the same system it is trying to protect. The malware detector performs its protection through the manifested malware detection technique(s), and serves as an empirical means of evaluating malware detection techniques' detection capabilities.

The rest of the paper is organized as follows. Section 2 discuss about the related works based on propagation of malware. In section 3, the proposed system has been described which includes system architecture and detailed description of each stage of the proposed system. Section 4 describes the system architecture. Section 5 summarizes the contents of this paper.

II. Related Works

M. E. J. Newman, Stephanie Forrest, and Justin Balthrop, [5] in this paper they work on to present an empirical analysis of the networks over which computer viruses spread and study some possible control strategies for preventing virus infections. Viruses typically arrive on a computer as an attachment to an Email message, which, when activated by the user, sends further copies of itself to other recipients. The Email addresses of these other recipients are usually obtained by examining an Email ‘‘address book,’’ a file in which the user for convenience stores the Email addresses of his or her regular correspondents. An important property of our Email network is that it is directed. That is, each edge (i.e., line) joining two vertices in the network has a direction. The directed nature of the network makes the spread of Email viruses qualitatively different from the spread of human diseases, for which most types of disease-causing contacts are undirected. Bidirectional edges can be thought of as undirected, and the Email network can be thought of as a ‘‘semi directed network,’’ a graph in which some edges are directed and others are undirected.

Shin-Ming Cheng, Member, IEEE, Weng Chon Ao, Pin-Yu Chen, and Kwang-Cheng Chen, Fellow, IEEE, [6] in this paper they work on to a novel differential equation-based model to analyze the mixed behaviours of delocalized infection and ripple based propagation for the hybrid malware in generalized social networks consisting of personal and spatial social relations. The malware on handsets typically exploits messaging services or uses SRWC services to propagate. The differential equation-based approach characterizing virus spreading in Internet is feasible to model the messaging malware dissemination due to homogeneity holds in person social network. On the other hand, the behaviour of malware spreading by SRWC services was approximated by differential equation or investigated by agent-based simulation. This paper proposes a novel analytical model to efficiently analyze the speed and severity for spreading the hybrid malware that targets multimedia messaging service (MMS) and BT. Validation against conducted simulation experiments reveals that our model developed from the Susceptible-Infected (SI) model in epidemiology accurately approximates mixed spreading behaviors in large areas without the huge computational cost, which helps estimate the damages caused by the hybrid malware and aids in the development of detection and containment processes. The proposed model is originated from the SI model in epidemic theory to measure propagation of infections within a population under risk. The communication between a compromised and a non compromised handset is modelled as a contact between an infected individual and a susceptible one, where a susceptible node acquires infection and never becomes susceptible again. This is due to the users' lack of concern about the threat of malwares and the limited capability of current antiviral software.

Cliff C. Zou, Don Towsley, and Weibo Gong, [3] in this paper they work on to study the topological impact, so compare Email worm propagation on power law topology with worm propagation on two other topologies: small world topology and random graph topology. The differential equation models presented by others cannot accurately model an epidemic spreading in a topological graph. Therefore, in this paper we will rely on simulation modeling rather than mathematical analysis in order to focus on realistic scenarios of Email worm propagation. The topology of an Email network plays a critical role in determining the propagation dynamics of an Email worm. Therefore, before start to study Email worm propagation, need to first determine the Email topology. One very important fact of an Email network (in terms of Email worm propagation) is that once a computer contains the address of an Email list, from an Email worm's point of view, this computer has virtually all the addresses associated with the Email list. Therefore, even though a user's computer may only contain tens of Email addresses, the degree of the user in the Email network might be as large as several thousand if one of the Email addresses is a popular Email list. For this reason, first study the property of Email lists. There are two major classes of epidemic models, defined by whether infected hosts can become susceptible again after recovery. If this is true, the models are called SIS models because hosts can change their status as susceptible-infectious-susceptible. If infected hosts cannot become susceptible again once they are cured, the models are called SIR models, hosts can only have the status transition as susceptible-infectious-recovered (or SI models if no infected hosts can recover). Our major focus in this paper is to understand the propagation dynamics of Email worms.

Existing investigations of malware propagation focus mostly on modeling the spread of malwares employing random scanning scheme. Random scanning selects targets to infect randomly. Malwares, however, can use other scanning methods. Although only a few topological malwares are known, topological scanning is a potential thread to the network routing infrastructure, World Wide Web (WWW) networks, and peer-to-peer systems, where topologies play an important role for malware propagation. For instance, contact process is used to analyze the ease of propagation on different topologies. The difficulty lies in characterizing the impact of topology and the interactions among nodes in both space and time. Such interactions result in a complex spatial-temporal dependence, which is especially hard to model.

Existing works presented certain strategies to immunize a group of users in the network to prevent topological worms from propagating to a large scale. However, how to choose the appropriate size and membership of this subset to constrain topological worm spreading remains a difficult question. A common view for the preferable positions of defense is at the highly-connected users or those with most active neighbors. Indeed, popular users in a scale-free network and their intuitively short paths to other nodes in a strongly clustered small world greatly facilitate the propagation of an infection over t the whole network, particularly at their early stage.

III. Proposed System

The existing analytical model [3] presented the spreading procedure by a susceptible-infected-susceptible (SIS) process, while it does not consider the new features of modern email malware. These observations become the motivation of our work to develop a new analytical model that can precisely present the propagation dynamics of the modern email malware. Since the spreading procedure can be characterized by a susceptible-infected-immunized (SII) process, so the proposed model is named as SII. SII model is different from SIS and SIR models because both susceptible and infected users can be immunized and never become susceptible again.SII model are used to overcome the inaccuracy of the existing model. In this proposed system, it is assumed that states of neighbouring nodes are independent. Nodes and topology are the basic elements for the propagation of modern email malware.

To obtain the email malware propagation mechanism following four steps are performed:

(i) virtual node generation
(ii) SII model
(iii) Propagation Dynamics
(iv) Email Recovery

VIRTUAL NODE GENERATION

Nodes and topology information are the basic elements for the propagation of modern email malware. Electronic mail allows a user to send a message to one or more recipients. A mail system allows a message to be send to multiple recipients. For modern email malware, a compromised user may send out malware email copies to neighbours every time the user visits those malware hyperlinks or attachments. Malware emails are also sent out when certain events like computer restart are triggered. At an arbitrary time t, a user may receive multiple malware email copies from an identical neighbouring user who has been compromised. In order to represent the repetitious spreading process of the reinfection and the self-start, introduce virtual nodes to represent the k^{th} infection caused by infected users opening the k^{th} malware copy.

SII Model

A node in the topology represents a user in the email network [7]. Let random variable X_i(t) denote the state of a node i at discrete time t. Then,

$$X_i(t) = \begin{cases} \textbf{Hea.}, \textit{healthy} \begin{cases} \textbf{Sus.}, & \textit{susceptible} \\ \textbf{Imm.}, & \textit{immunized} \end{cases} \\ \textbf{Inf.}, \textit{infected} \begin{cases} \textbf{Act.}, & \textit{active} \\ \textbf{Dor.}, & \textit{dormant.} \end{cases} \end{cases}$$

In SII Model, derive an M by M square matrix with elements pij to describe a topology consisting of M nodes, as in

$$\begin{pmatrix} p_{11} & \cdots & p_{1M} \\ \vdots & p_{ij} & \vdots \\ p_{M1} & \cdots & p_{MM} \end{pmatrix} p_{ij} \in [0, 1],$$

Where in pij represents the probability of user j visiting a deceptive malware email received from user i. If pij is equal to zero, it means the email address of user j is not in the contact list of user i. Therefore, the matrix reflects the topology of an email network. In this model, it is assumed that states of neighbouring nodes are independent. The infection of email malware depends on unwary email users checking new emails and visiting those malicious ones. An email user may receive multiple emails at different time, but read all of them at one time when the user checks the mailbox.

PROPAGATION DYNAMICS

To represent the spreading process of virtual nodes, extend Ni (the set of neighbouring nodes of node i) into a new set of neighbouring nodes, Hi, which contains three subsets: Ni/N, Ni/R and Ni/S.

- First, the subset Ni/N includes the real neighbouring nodes of user i. The nodes in Ni/N represent the neighbouring users who visit the first malware email copy and get infected.
- Second, the subset Ni/S includes the virtual nodes which present the extra spreading processes caused by certain events triggered in infected nodes.
- Third, the subset Ni/R includes the virtual nodes which present the extra spreading processes caused by users visiting more than one malware copies when they check new emails.

There are three preconditions for an arbitrary user being infected by email malware:

1. The user has not been immunized.
2. The user checks mailbox for new emails.
3. The user unwarily visits one received malware emails.

EMAIL RECOVERY

For an email network, immunizing a node means that the node can’t be infected by the email worm under study. In this module, consider a static immunization defense. An email worm starts to propagate a small number of nodes in the network have already been immunized. If some email users are well educated and never open suspicious email attachments, they can be treated as immunized nodes in the email network.

IV. SYSTEM ARCHITECTURE

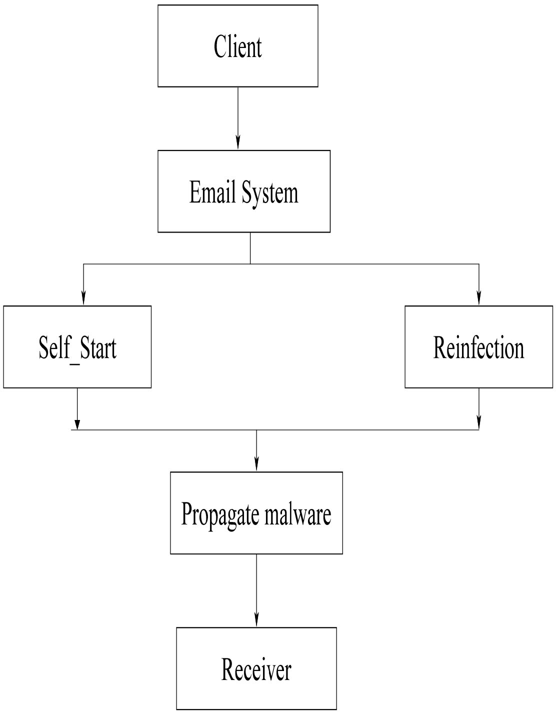

Figure 4.1: System Architectural Diagram Representation

For modern email malware, recall that a compromised user may send out malware email copies to neighbours every time the user visits those malware hyperlinks or attachments. Malware emails are also sent out when certain events like computer restart are triggered. Modern email malware exhibits two new features, reinfection and self-start. Reinfection refers to the malware behaviour that modern email malware sends out malware copies whenever any healthy or infected recipients open the malicious attachment. Self-start refers to the behaviour that malware starts to spread whenever compromised computers restart or certain files are visited. Reinfection, as the name suggests, indicates a user may get infected whenever the user visits malicious hyperlink or attachments.

The malware is propgated using two mechanisms; reinfection and self start. The client can send email to aother clients. The email messages are received by the srever. The server checks the emails and forwareded to the recepients.

V. CONCLUSION

In this paper, derive a new novel susceptible-infected-immunized model for the propagation dynamics of modern email malware. This model can address the two critical problems of modern email malware: self-start and reinfection. By introducing a group of difference equations and virtual nodes, we presented the repetitious spreading processes caused by the reinfection and the self-start. For the future work, there are also some problems needed to be solved, such as independent assumption between users in the network and in the malware detection techniques. In this paper, mainly focus on presenting the reinfection and the self-start in the modelling.

ACKNOWLEDGMENT

It gives me immense pleasure to express my gratitude to all those who encouraged me throughout the thesis work. First of all I would like to thank God Almighty for all the blessings he has showered on me always.I express my sincere thanks and a deep sense of humble gratitude to the Principal of this Institution, Dr. M. K. Jana for providing us all necessary facilities.I wish to express my sincere gratitude and respect to Dr.C.G.Sukumaran Nair, Head of the Department, Computer Science and Engineering, for all the words of inspiration.I wish to express my sincere gratitude and respect to Sudha.S.K, Associate professor, Computer Science and Engineering, for all the words of inspiration.I would like to express my profound gratitude to my project guide Mrs.SURYAPRIYA S, Assistant Professor for her invaluable support, encouragement, supervision and useful suggestions throughout this work.I take this opportunity to thank my parents and friends but no words to express my gratitude towards them who always stayed with me as my support and encourage.

REFERENCES

[1]. Palme, Karlgren & Pargman (1994): Issues when designing filters in messaging systems. Computer Communications, vol. 19, no 2, pp 95-101.

[2]. Z. Chen and C. Ji, “Spatial-Temporal Modeling of Malware Propagation in Networks,” IEEE Trans. Neural Networks, vol. 16, no. 5, pp. 1291-1303, Sept. 2005.

[3]. 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.

[4]. Mohsen Damshenas, Ali Dehghantanha, Ramlan Mahmoud,” A Survey on Malware Propagation, Analysis, and Detection”, International Journal of Cyber-Security and Digital Forensics (IJCSDF) 2(4): 10-29, the Society of Digital Information and Wireless Communications, 2013 (ISSN: 2305-0012).

[5]. M.E. Newman, S. Forrest, and J. Balthrop, “Email Networks and the Spread of Computer Viruses,” Physical Rev. E, vol. 66, no. 3, 2002.

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

[7]. G. Serazzi and S. Zanero, “Computer Virus Propagation Models,” Proc. 11th IEEE/ACM Int’l Conf. Modeling, Analysis and Simulations of Computer and Telecomm. Systems (MASCOTS ’03), pp. 1-10, Oct. 2003.

[8]. G. Yan, G. Chen, S. Eidenbenz, and N. Li, “Malware Propagation in Online Social Networks: Nature, Dynamics, and Defense Implications,” Proc. Sixth ACM Symp. Information, Computer and Comm. Security (ASIACCS ’11), pp. 196-206, 2011.

[9]. H. Ebel, L.I. Mielsch, and S. Bornholdt, “Scale-Free Topology of Email Networks,” Physical Rev. E, vol. 66, no. 3, Sept. 2002.