ISSN 2348-375X

Unique Journal of Engineering and Advanced Sciences

Available online: www.ujconline.net

Research Article

MODELING AND STUDY ON THE PROLIFERATION BEHAVIORS OF RECENT EMAIL SPAMWARE

Vinoth Kumar J[1*], Vikramarajan Jambulingam[2]

[1]Department of Computer Science and Engineering, Bhajarang Engineering College, Tamil nadu, India
[2]Department of Electrical and Computer engineering, University of Gondar

Received: 28-04-2015; Revised: 27-05-2015; Accepted: 26-06-2015

*Corresponding Author: **J. Vinoth Kumar**
Department of Computer Science and Engineering, Bhajarang Engineering College, Tamil nadu, India, Mobile: 09677823604

ABSTRACT

Due to the serious security threats forced by email-based malware in latest years, modeling the proliferation methods of email malware becomes an essential system for predicting its possible costs and creating efficient countermeasures. When compared to former occurrences of email malware, current email malware portrays two new characteristics, reinjection and self-start. Reinjection means the malware activity that sends away malware copies each time any vigorous or infected recipients release the malicious attachment. Self-start means the activity that spreads each time compromised systems restart or certain files are opened. In the literature, there were many models that exists for email malware proliferation, but they did not include the above two characteristics and does not correctly model the dissemination dynamics of modern email malware. To solve this issue, we obtain a new difference equation based logical model by introducing a new concept of near infected user. The proposed model can accurately near the repetitious distribution process caused by reinjection and self-start and successfully overcome the related computational complexities. We perform complete experimental and notional study to confirm the proposed logical model. The outputs demonstrate our model vastly performs better than the previous models in terms of correctness.

Keywords: Information security, Reinjection, email malware, proliferation modeling.

INTRODUCTION

Electronic mail is basically a service for computer users but malware that comes with some email poses significant security threats. For several years, the propagation of email based malware has been following the same methodology. A malware email if sent to the victim will appear as though it was sent by somebody the recipient believes in. The subject is also somewhat related to the recipient's area of interest. Once the victim is faked into either clicking the malicious hyperlinks or accessing the attachments inside an email, the computer will be compromised immediately. Then, the compromised system will start infecting new tar-gets that are found in its email address lists. To stop email malware, researchers have put efforts to deter people from accessing unwanted hyperlinks and email attachments. But, the success of current new email malware like "Here you are"[1], indicates that these educational measures are not enough. Previous works[2-6], explain that a user can get infected and send out malware copies only once, whether or not the user opens a malicious hyperlink or attachment again. Some examples are those early email malwares such as Melissa in 1999[7], which will check whether a victim has been already compromised or not before the infection. But, modern email malware is far more lethal to disseminate in network than ever before by introducing two new characteristics[5],[9],[10]. First feature is "reinjection", i.e., an infected user sends out malware copies multiple times this user visits the malicious hyperlinks or attachments. Second feature is "self-start", i.e., an infected user sends out malware copies when certain situations like PC restart occur. Researchers stated that a user can be infected multiple times.

It is a big challenge to explore current email mal-ware through mathematical modeling. The previous analytical model presented the spreading process by a susceptible-infected-susceptible (SIS) process but it does not consider the new features of current e-mail malware. These comments become the inspiration of our work to develop a new analytical model that can exactly present the proliferation dynamics of the recent email malware.

The major offerings of this work are listed below: We propose

a new logical model to detain the connections among the infected email users by a set of difference equations, which together describe largely propagation of the current email malware. We do experimental and notional study to investigate why and how the propose model is better to existing models.

The rest of the paper is prepared as follows. Section 2 states the problems in modeling current email malware. In Section 3, a new analytical model is presented. Section 4 details a series of experiments to check the proposed model. Further conversation and related work are offered in Sections 5. Finally, Section 6 concludes this work.

PROBLEM STATEMENT

Selecting email as the spreading carrier of malware is an old technique in the last decade. Usually compromised user will send out malware emails only once, after which the user may not send out any further copies of malware, even if he visits the malicious hyperlinks or attachments again. If you take Melissa for example, the malware first checks a precise registry key in the Window OS and the malware will not do anything further when the value of this key suggests that the user has already been infected before. In the following, we call this kind of spreading mechanism as non-reinjection. Without examination if a computer has been infected before, recent email malware will take every chance to spread itself. We exemplify its propagation with two kinds of new methods namely reinjection and self-start.

Problem from practical Perspective

Reinjection, as the name itself indicates a user may get infected whenever the user opens malicious hyperlink or attachments. The reinjection negates the non-reinjection in two aspects: 1) A user can be infected again even if the user has been infected before. 2) A user will send out a malware copy every time the user gets tainted. Thus, a recipient may frequently obtain malware emails from the same compromised user.

In case 1 of the non-reinjection, even a user i opens two malware emails at t_8, the user will get infected and sends only one copy of malware to user j at t_8. The malware email comes at user j at t_9. When user j checks inbox at t_{13} and opens the malware email from user i, user j gets infected. User j will not get any more malware emails from user i after t_9. Nevertheless, in case 3 of the reinjection, user j will get two malware copies from user i at t_9. Then, after user j gets infected at t_{13}, when user i opens another two malware emails, user j receives another two copies of malware from user iatt$_{17}$. Compared with case 1 of the non-reinjection, user j in case 3 of the reinjection gets totally four malware emails.

In Table 1: We list some types of email malware

Name	Subject	Message	Attachment	Hyperlink
Sircam	random	random	random	none
SoBig	13	2	13	none
Mydoom	19	8	26	none
Nyxem	23	8	36	none
NetSky	1	1	25	none
W32.Imsol	2	7	none	none

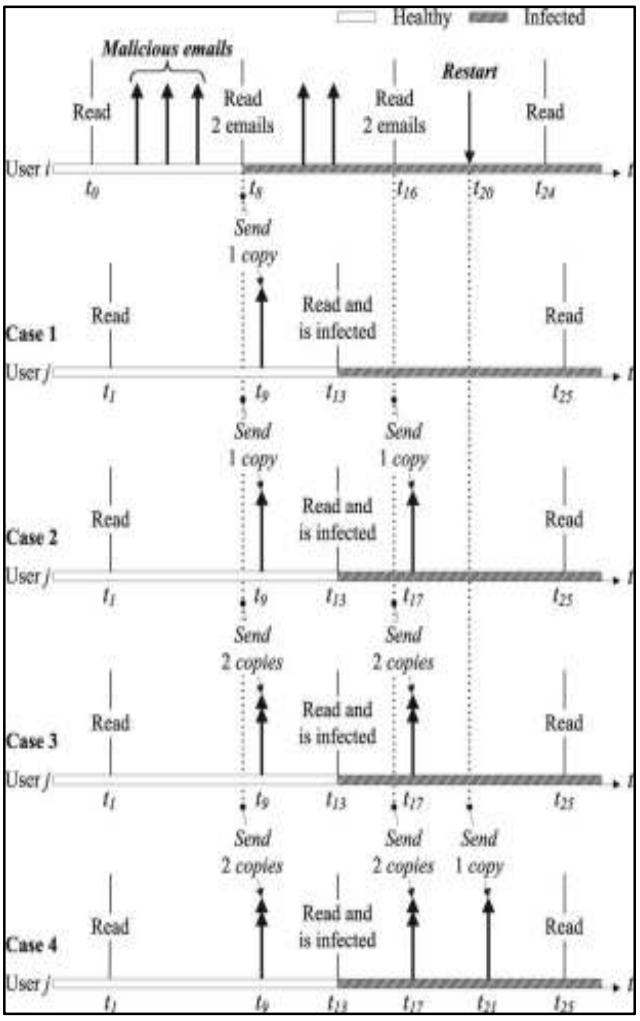

Figure 1: Recipient user j's pattern for various types of malware emails. User i open two of three malware emails at t_8 and another two at t_{16}, and then restart at t_{20}. Case 1: non-reinjection; Case 2: reinjection in the work; Case 3: reinjection of current email malware; Case 4: both self-start mechanism in current email malware. We guess a user will call the malicious hyperlink or attachment if the user opens emails in this figure.

Results from Symantec Security Response

In this paper, we assume every malicious email has different themes. In fact, reinjection is also not enough to describe the proliferation of current email malware. In many cases, they change registry entries in Windows OS and the spreading process can be caused whenever compromised systems restart or certain files are opened by affected users. For example, take my doom it runs each time Windows starts.

Problem from experimental Perspective: Currently, several models have been offered to model the proliferation of email malware. For example, the earlier works present the non-reinfection[2-5]; the other earlier works[10], model the reinjection and the past work [5] also explains the self-start. First, can we use the models of the non-reinjection to present the proliferation dynamics of current email malware? Compared with the reinjection and the self-start to model the no reinjection is simple. But the self-start mechanism can spread much quicker.

Second, can we use the earlier models of the reinjection and the self-start to describe the proliferation of current e-mail

malware? The differential equation model adopted in [10] has been proven by the past work to overrate the dispersal speed by 20 percent. Because the work [5] does not give enough details in modeling the reinjection and the self-start, we mainly show to the earlier works.

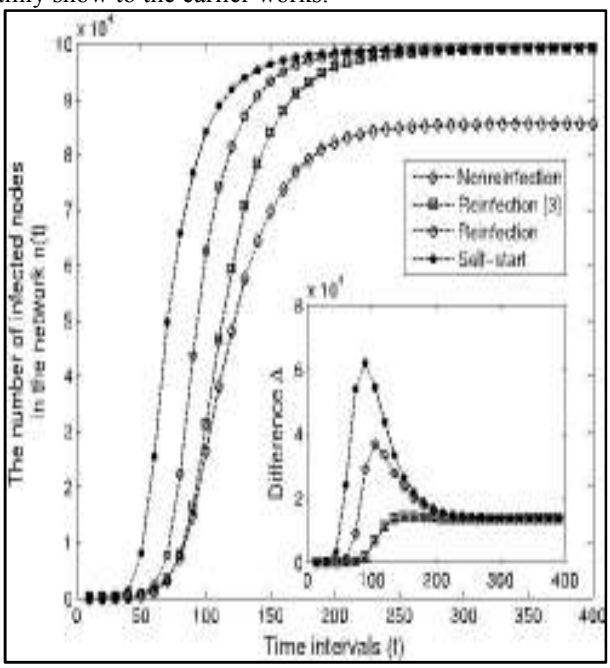

Figure 2: The proliferation of email malware in a network with 10^6 users. The results are a mean from 100 simulations. The figure gives the differences (D) of many spreading mechanisms to the no reinjection mechanism. So, the earlier models of the reinjection and the self-start shall not be used in the proliferation of current email malware.

SII MODEL: In order to overcome the incorrectness of earlier models, we expand our earlier SII model[6] for current email malware. SII model is different from SIS and SIR models because both vulnerable and tainted users can be immunized and never become vulnerable again.

Modeling Nodes, Topology and Events of User: Nodes and topology information are the basic elements for the proliferation of recent email malware. A node in the topology indicates a user in the email network. Let random variable X_i (t) denote the state of a node i at discrete time t.

Then, we have

$X_i(t)$ = Hea.. Healthly (either Sus – susceptible or Imm - Immunized)

$X_i(t)$ = Inf.. Infected (either Act – active or Dor - Dormant)

The state transition graph of a random node i in an e-mail network is drawn in Fig. 3. All nodes in networks are initially assumed to be in a susceptible state. Since infected users will send out copies of malware if they are compromised, node i move from the susceptible state to the active state after the node i's user gets infected. The infection odds is denoted by v(i,t). The user is infectious at the active state. When a user is infected but not infectious, then the user node moves to the dormant state. Besides, any user node can be compromised again even if the user has been already infected before. We represent the infection odds of an arbitrary node being at the dormant state and active state as g(i,t) and h(i,t) correspondingly.

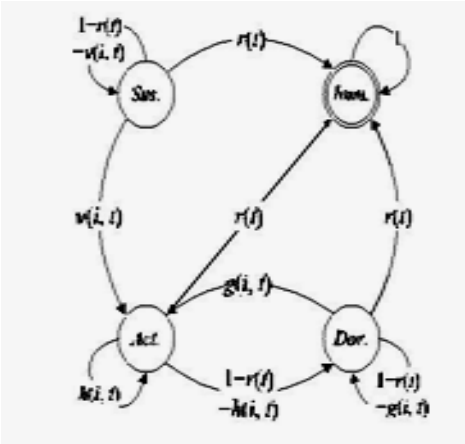

Figure 3: State transition graph of a node in email network. "Sus.": healthy but susceptible; "Act.": a user is infectious and will send copies to infect others; "Dor.": a user is injected but not yet infectious; "Imm.": healthy and will not be injected again.

Whatever the state an random node is at, it may move to the immunized state. The probability of immunization is represented by r(t). But if the values of g(i,t) and h(i,t) are equal to zero, any infected node i will be present at the dormant state till the user of that node is immunized. In our proposed model, we propose using an M by M square matrix with elements p_{ij} to describe a topology containing of M nodes as follows,

$$\begin{bmatrix} p11 & \dots & p1m \\ . & pij & . \\ pm1 & \dots & pmm \end{bmatrix} p_{ij} \in [0,1]$$

Wherein p_{ij} represents the possibility of user j opening a misleading malware email received from user i. If p_{ij} is equal to zero, it indicates the email address of user j is not in the contact list of user i. In this p, we take the states of neighboring nodes as independent.

Modeling Propagation Dynamics: we use the values 0 & 1 to replace with the healthy state and then obviously for the infected state. Given a structure of an email network with M nodes, the expected number of infected users at time t, is computed as in

$$n(t)=E\left[\sum_{i=1}^{M} P(X_{i(t)} = Inf)\right]$$

The Expected number of infected nodes n(t) is assigned to the sum of the probability of each node getting infected at time t, P(Xt (t) = Inf) As seen in fig 3. A susceptible node must be compromised and be at the infected state and an infected node can be recovered and be at the immunized state.

$$P(X_i(t) = Inf.) = (1 - r(t).P(X_i(t-1) = Inf.)$$
$$+v(i,t).P(X_i(t-1) = Sus.)$$

Virtual Nodes: For current email malware, remember that a compromised user can send out malware email copies to neighbors each time the user visits those malware hyperlinks or attachments. Malware emails are also sent out if specific events like computer restart are caused. Thus, at an random time t, a user may get multiple malware email copies from an

identical neighboring user who has been compromised. In order to represent the repetitious distribution process of the reinjection and the self-start, we introduce virtual nodes to present the nth infection caused by infected users opening the nth malware email copy. As shown in Fig. 4, node 1, 2, 3 send malware emails to node 4. When the user of node 4 visits those emails, the user gets infected. If the user of node 4 visits two malware emails, node 4 will send malware email copies twice to node 6. If the user of node 4 visits three malware emails, node 4 will send treble malware email copies to node 6.

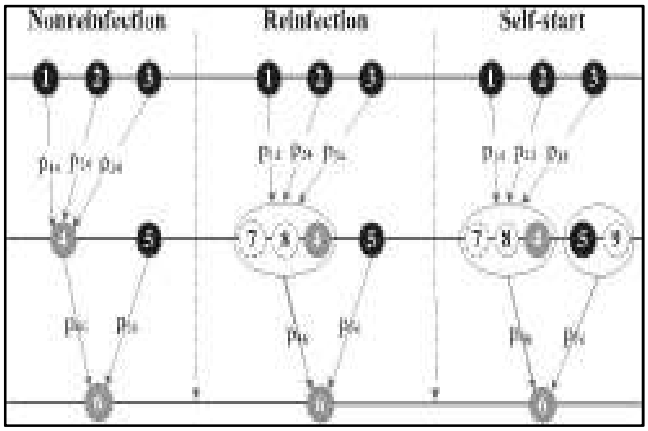

Figure 4: An example to show virtual nodes in the reinjection and slow start phase

MODEL EVALUATION

Experiment Environment:

In this field, all existing research adopts simulation to evaluate analytical models, such as [2, 4]. we follow this approach to evaluate the proposed SII model based on simulations. In real-world scenarios, the spread of most email malware is typically impossible to track given the directed, topological manner in which they spread. It should be pointed out that there is no real data set avail-able for the evaluation of models of modern email malware.

Comparison with Previous Models:

To evaluate the accuracy of our model, we conduct experiments with different parameter settings. Most of the values, we can compare our work with earlier works but those models do not present virtual nodes and reinjection. We can see that the results of previous models deviate from simulations by 80 thousands less infections at maximum. There is also a minor divergence between the results of SII model and simulations.

Impact of Parameters in the Modeling:

We also evaluate the impact of various parameters on the correctness of our modeling. First, we evaluate the accuracy with different distributions of T_i and R_i. In this experiment, the topology has the same settings as in Fig. 7. As shown in Fig. 5, the curves of our SII model are close to the simulations even if the distributions of T_i and R_i are different. Second, we also evaluate the accuracy with different distributions of p_{ij}. The same topologies are used in this experiment. We let T_i and R_i follow Gaussian distribution. As shown in Fig. 5, the results of our SII model are close to the results of simulations. In the inset figure of Fig. 5, we can also see that the SII model achieves better performance in accuracy when the infection probabilities p_{ij} are averagely higher. For the same reason of the independent assumption, we can achieve better accuracy once we relax this assumption in the future modeling.

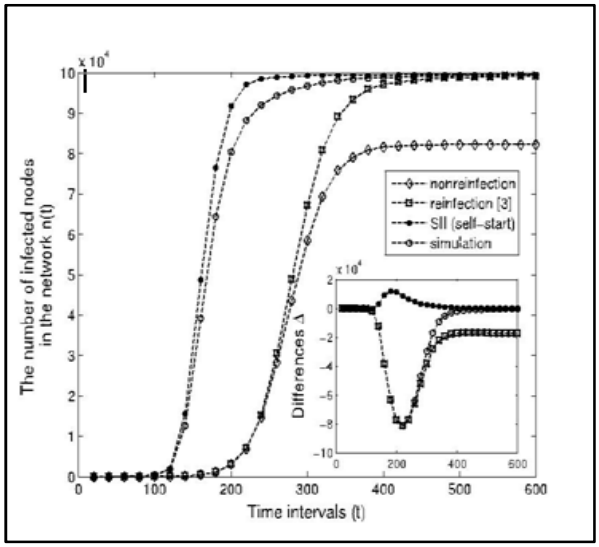

Figure 5: The comparison between SII model and other models

RELATED WORK

There have been substantial efforts in modeling the propagation dynamics of Internet malware in the last decade. First, to model the epidemic spreading on topological networks, early researchers adopt differential equations to present the propagation dynamics of malware. However, the differential models greatly overestimate the spreading speed due to the "homogeneous mixing" assumption. Additionally, ZouandGao rely on simulations to model the spread of email malware. Their simulation models avoid the "homogeneous mixing" problem but cannot provide analytical propagation studies. The works [2,6] propose mathematical models, which have captured the accurate topological information. Wen et al. [6]further addressed the temporal dynamics and the spatial dependence problem in the propagation model-ling. However, all these models cannot present the reinfection and self-start processes of modern email malware.

The works in focus on threshold conditions for malware fast extinction on the Internet. Their works study the final stable state of epidemic spread based on SIS models, whereas we study the transient propagation dynamics of modern email malware[12].Second, there are some works which characterize the propagation dynamics of isomorphic malware, such as P2P malware, mobile malware, [15] and malware on online social networks[13,14]. R. Thommes and M. Coates adopt differential equations to present the propagation of P2P malware through a P2P network. The models are proposed for the mobile environment by presuming nodes meet each other with a probability[15]. These works assume all individual devices are homogeneously mixed, and thus, they are unlikely to work in the real mobile environment. The models present the propagation of online social malware. Since these models are based on non-reinjection; they cannot be adopted to present the propagation of modern e-mail malware[13-15].

CONCLUSION

In this paper, we have proposed a new SII model for the proliferation of recent email malware. This model addresses two critical processes not solved in earlier models: the reinjection and the self-start. By developing a group of difference equations and virtual nodes, we presented the automatic spreading processes caused by the reinjection and the self-start. The experiments revealed that the result of our model better. For the further work, there are also some problems that are to be solved, such as the independent assumption between users in the network and the periodic assumption of email checking time of users.

REFERENCES

1. Fossi M and Blackbird J, Symantec Internet Security Threat Report 2010, technical report Symantec Corporation, 2011.
2. Chen Z and Ji C, Spatial-Temporal Modeling of Malware Propagation in Networks, IEEE Trans. Neural Networks, 2005; 16(5):1291-1303.
3. Gao C, Liu J, and Zhong N, Network Immunization and Virus Propagation in Email Networks: Experimental Evaluation and Analysis, Knowledge and Information Systems, 2011; 27(2):253-279.
4. Wen S, Zhou W, Wang Y, Zhou W, and Xiang Y, Locating Defense Positions for Thwarting the Propagation of Topological Worms, IEEE Comm. Letters, 2012;16(4) : 560-563.
5. Xiong J, Act: Attachment Chain Tracing Scheme for Email Virus Detection and Control, Proc. ACM Workshop Rapid Malcode, 2004:11-22.
6. Wen S, Zhou W, Zhang J, Xiang Y, Zhou W, and Jia W, Modeling Propagation Dynamics of Social Network Worms, IEEE Trans. Parallel and Distributed Systems,2013; 24(8) : 1633-1643.
7. Cert, advisory ca-1999-04, Melissa Macro Virus, http:// www.cert.org/advisories/CA-1999 04.html, 2009.
8. Cert, Advisory ca-2000-04, Love Letter Worm, http://www.cert. org/advisories/CA-2000-04.html, 2000.
9. Calzarossa M and Gelenbe E, Performance Tools and Applications to Networked Systems: Revised Tutorial Lectures. Springer-Verlag, 2004.
10. Rozenberg B, Gudes E, and Elovici Y, SISR: A New Model for Epidemic Spreading of Electronic Threats, 2009; 242-249.
11. Cert, Advisory ca-2001-22, w32/sircam Malicious Code, http://www.cert.org/advisories/CA-2001-22.html, 2001.
12. Ganesh AJ, Massouli L, and Towsley D F, The Effect of Network Topology on the Spread of Epidemics, Proc. IEEsE INFO-COM ,2005:1455-1466.
13. Yan G, Chen G, Eidenbenz S, and Li N, Malware Propagation in Online Social Networks: Nature, Dynamics, and Defense Implications, Proc. Sixth ACM Symp. Information, Computer and Comm. Security 2011:196-206.
14. Fan W and Yeung KH, Online Social Networks-Paradise of Computer Viruses,Physica A: Statistical Mechanics and Its Applications, 2011; 390(2):189-197.
15. Cheng S M, Ao W C, Chen P Y, and Chen K C, On Modeling Malware Propagation in Generalized Social Networks, IEEE Comm. Letters, 2011; 15(1): 25-27.

J. Vinothkumar received his Master and Bachelor degree in Computer Science Engineering from affiliated colleges of Anna University, India. His research interests are computer networks, network security and cloud computing.

Vikramarajan Jambulingam received his Master degree in Power Electronics and Drives and Bachelor degree in Electrical and Electronics Engineering from VIT University, India. His research interests are power electronic applications, power quality, power electronic converters and computer networks.

Source of support: Nil, Conflict of interest: None Declared