ISSN: 0976-3104
SPECIAL ISSUE: (Emerging Technologies in Networking and Security (ETNS)

THE IIOAB JOURNAL

ARTICLE OPEN ACCESS

ANALYZE AND PREVENT MODERN EMAIL MALWARE PROPAGATION USING SEII MODEL

S. Sneha*, P. Swapna
Dept of Information Technology, M. Kumarasamy college of Engineering, Thalavapalayam, Karur-639113, Tamilnadu, INDIA

ABSTRACT

Aim: *It is necessary to develop a prevention model to avoid potential damages caused by modern email malware. While comparing to earlier versions the modern email malware has two characteristics: reinfection and self-start. In earlier research some models were developed for analyzing the malware propagation still there is needed to improve the accuracy of the model. Further the existing approach uses virtual node concept which increases the computational overhead. To concentrate on these problems, SEII model is proposed to analyze and prevent the modern email malware. Based on the result of the analysis model the impact of parameters in propagation is analyzed and presents the automated email malware detection and control system. Complete inspection and demonstration shows that the proposed model can detect and prevent the modern email malware in effective manner.*

Published on: 2nd -December-2016

KEY WORDS

email malware, reinfection, self-start, SEII model, control system

*** Email:** snekasekaran@gmail.com; **Tel.:** +91 7871473343

INTRODUCTION

The email malware creates the major security issues for email service users. A computer virus is one of the major forms of malicious information spreading in the Internet. The computer viruses are classified into scanning-based viruses and topological-based viruses. The email malware is based on the topological viruses. Once an email user is infected by email malware, malicious email copies are sent to their friends embedded in email lists. The user will get infected, whenever they open and read the email received from the infected user. The infection processes are spreads quickly and reach a large scale, frequently from one user to adjacent users.

Current research on email malware focuses on propagation dynamics [1, 2, 3, 4, 5] which is used to decrease email malware distribution speed. In email malware, if the user visits a malicious mail again or not the infected user will send the email copies once[1, 2, 3, 6]. The modern email malware is extremely dangerous because of reinfection and self-start. In reinfection, every time healthy or infected recipients open the malicious attached file the modern email malware sends its copy to all users in list. In self-start, each time compromised computers restart or malicious files are visited the malware, and then the malware occupied the memory of the system by spreading its own copy. The existing analytical SII model[10] presented the procedure of malware spreading, but it cannot accurately estimate the spreading of email malware. A SEII systematic model is proposed to describe the email malware propagation. The distribution procedure can be distinguished by a susceptible-exposed-infected-immunized (SEII) process.

In this paper, we define virus transmission rules as follows:

- If susceptible nodes get contact with an infectious node, then it transmits into an exposed.
- If exposed nodes get contact with an infectious node, then that will transmit into an infected.
- All the nodes transmits to an immunized state, when the user immunized.

[13] proposed an analytical model that[13] represented the spreading process by susceptible-infected-susceptible (SIS) process. In this model, susceptible and infected users both can be susceptible again. In [1] [18] presented the SIR model to describe the email malware propagation. In

this model both susceptible and infected users can be recovered and they would receive lifelong immunity. The work of [6][9] characterized the propagation dynamics of isomorphic malware, such as P2P malware[1], mobile malware[14, 15] and malware on online social networks [16, 15, 17]. These existing models are not able to effectively present the propagation of modern email malware.

To tackle the problem of reinfection and self-start Sheng Wen[10], proposed a novel analytical SII model. It cannot perfectly estimate the spreading of modern email malware. This model had some minor deviation between mathematical model and simulations results because of the independent assumption.

PROBLEM DEFINITION

The malware propagation is based on non-reinfection, refection and self-start. In fact, reinfection alone is not sufficient to describe the propagation of modern email malware because most email malware is having the self-start. Modeling the non-reinfection is simple when compared to reinfection and self-start. Therefore, reinfection and self-start are the two mechanisms used to differentiate the spreading of modern email malware.

Reinfection denotes whenever the infected user opens the malicious mail he will get infected again. The reinfection dominates the non-reinfection in below aspects:

(i) User will infect again even if he/she infected before.
(ii) Whenever the user get infected he/she will send the mail to their neighbor
Thus, a receiver may continually receive malicious emails from the same compromised user.

Self-start is the behavior of spreading malware whenever the infected computers restart. The sur pass of the self-start is given below:

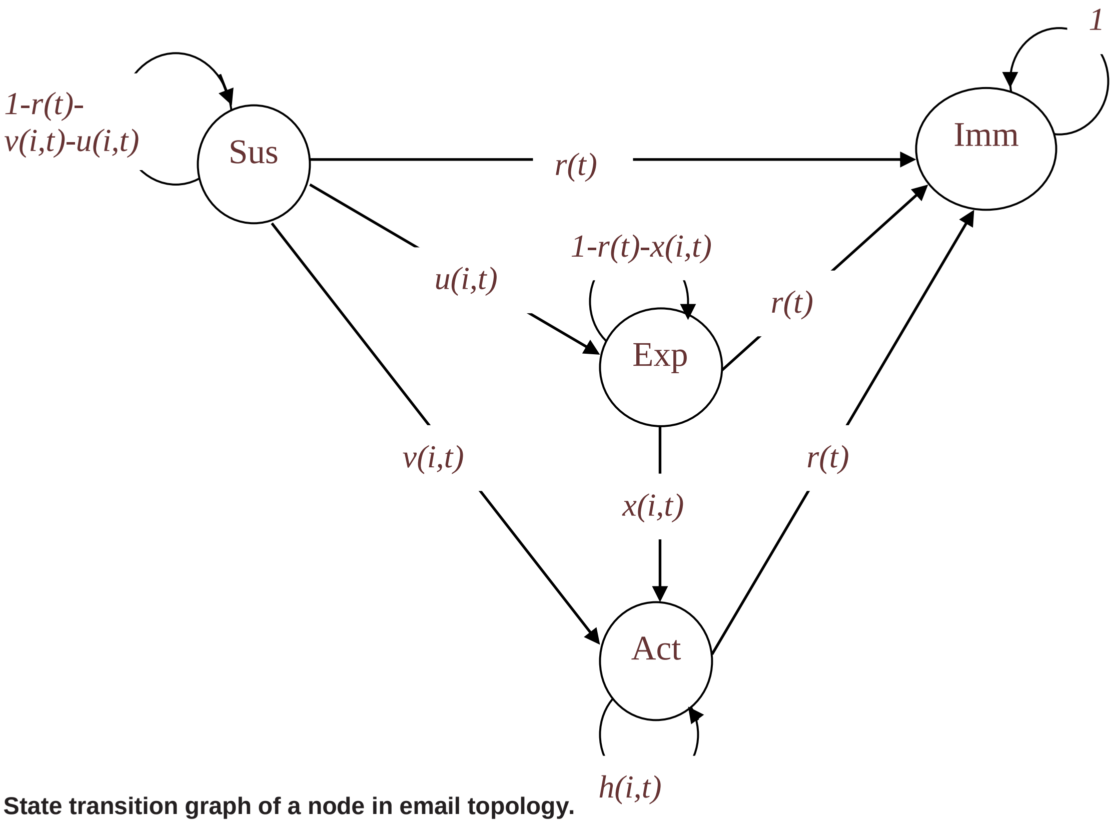

Fig: 1. State transition graph of a node in email topology.

THE IIOAB JOURNAL ISSN: 0976-3104

A user has been infected at a particular time. If the user restarts the computer, a malicious email copy is received by a new user because of self-start. The malware can spread much faster in self start, when compare to all other models.

To conquer these complexities, SEII analytical model is proposed which will describe the dissemination of the modern email malware.

PROPOSED SYSTEM

To overcome the difficulties of previous models, we propose SEII model as shown in **[Figure- 1]** for modern email malware. SEII model is accurate than SIS and SIR models [18, 10] and SII model[10]. In this model, susceptible, exposed and infected users can be immunized.

SEII MODEL - INTRODUCTION

The essential elements for the dissemination of modern email malware are topology and node information. A node represents a user in the email network. Let random variable Xi (t) indicate the state of a node i at distinct time t.

$$X_i(t) = \begin{cases} Hea.,Healthy \begin{cases} Sus.,Susceptible \\ Exp.,Exposed \\ Imm.,Immunized \end{cases} \\ Inf.,Infected \end{cases} \quad (1)$$

Initially all the nodes in the network are susceptible. In susceptible state the user have the possibility of getting infected. The susceptible node is transits to active state when the user infected. If the user i is in the address book of infected user, then the infection possibility of the user i is higher. Therefore, the user transmits from the susceptible state to the exposed state. Since, the infected user sends out the malware to the user i when it is compromised, then the user i transmits from exposed state to active state after the infection of user i. All the states are transmits to immunized state at final.

Let r (t) be the probability of immunization. H (i, t) is the probability of being in the active state. V (i, t) is the probability of node i transits from susceptible to active state. U (i, t) be the probability that node i transits from susceptible to exposed state. X (i, t) is the probability that the node i moved to active state from exposed state. In SEII Model, an N by N square matrix with elements pij is used to describe a network topology consisting of N nodes, as in,

$$\begin{pmatrix} p_{11} & \cdots & p_{1M} \\ \vdots & p_{ij} & \vdots \\ p_{M1} & \cdots & p_{MM} \end{pmatrix} p_{ij} \in [0,1] \qquad (2)$$

Where in pij stands for the probability, that the user j visits a malicious email from user i. If pij is equal to zero, mail address of user j is not present in the contact list of user i. Therefore, the matrix replicates the topology of an email network. This model assumes the states of neighbouring nodes are dependent.

The infection of malware depends on unsuspecting email users' read-through new emails. This process involves two components in the modeling. First, the flag variable openi(t) is introduced. Openi(t)=1, if the newly arrived mails are checked by user at time t, otherwise openi(t)=0.

$$P(open_i(t) = 1) = \begin{cases} 0\ , otherwise \\ 1, t\ mod\ T_i = 0 \end{cases} \quad (3)$$

Ti – user i's email checking period.

Every user has different values of Ti. User can receive numerous emails at different time but the user checks the mailbox and read the email at one time. Suppose the user i checks newly received emails at time t, then the user checks email at time t. That is user i receives the new email after the last checking act of her mailbox. Here we initiate a variable t to indicate random time between the user i's last email checking time and the current time t. Then the value of t is,

$$\begin{cases} t - T_i \le \tau < t, & if\ open_i(t) = 1 \\ t - (t\ mod\ T_i) \le \tau < t, & if\ not \end{cases} \qquad (4)$$

An infected user only propagates malware to the adjacent users in topological networks. For each user in email networks, we record and collect every new malicious email from adjacent users at time t, and at last attain the joint infection probability of each user who checks the malicious emails.

SPREADING ANALYSIS

Here, the values 0 and 1 are used to substitute the healthy state and the infected state, respectively. M denotes the nodes in email network topology, the number of infected users at time t and n(t), is computed as in,

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

$$= \sum_{i=1}^{M} [(0P(X_i(t) = 0) + (1P(X_i(t) = 1))] = \sum_{i=1}^{M} P(X_i(t) = 1)$$

$$= \sum_{i=1}^{M} P(X_i(t) = \mathrm{inf}) \qquad (5)$$

The probable number of infected nodes, n(t), is equal to the sum of the probability of each node being infected a time t, $P(X_i(t) = \mathrm{inf})$. As made known in **[Figure- 1]**, susceptible node and an exposed node can be infected and stay in the infected state, and an infected node recovered and stay at the immunized state. The transitions of each state is used to derive the $P(X_i(t) = \mathrm{inf})$ by difference equations as follows:

$$P(X_i(t) = Inf) = (1 - r(t))P(X_i(t-1) = Inf) + v(i,t)P(X_i(t-1) = Sus) + x(i,t)P(X_i(t-1) = Exp) \qquad (6)$$

To find $P(X_i(t) = Sus)$ we have,

$$P(X_i(t) = Sus) = 1 - P(X_i(t) = Inf) - P(X_i(t) = Exp) - P(X_i(t) = Imm) \qquad (7)$$

To find $P(X_i(t) = Exp)$ we have,

$$P(X_i(t) = Exp) = u(i,t)P(X_i(t) = Sus)_+ (1 - r(t))P(X_i(t-1) = Inf) + (1 - x(i,t))P(X_i(t-1) = Inf) \qquad (8)$$

To find $P(X_i(t) = Imm)$ we have,

$$P(X_i(t) = Imm) = P(X_i(t-1) = Imm) + r(t).[1 - P(X_i(t-1) = Imm)] \qquad (9)$$

Once the values of v (i, t), x(i,t) and r(t) are obtained then the value of $P(X_i(t) = Inf)$ can be calculated by using the iteration of the above equations, (6),(7),(8) and (9).

There are three prerequisites for random user being infected by email malware is given below:

i) the user should not immunized;

ii)the user checks for new emails;

iii) the user incautiously visits malicious emails;

When the first two prerequisites are satisfied, the s (i, t) is used to represent the probability of user i visiting malicious emails from adjacent nodes. Then, the probability for infection v (i, t) and x (i, t) can be derived as in,

$$v(i,t) = s(i,t).P(open_i(t) = 1)(1 - r(t)) \quad (10)$$

$$x(i,t) = s(i,t).P(open_i(t) = 1)(1 - r(t))\,(1 - u(i,t)) \quad (11)$$

In SEII model, an user i visits malicious attachments with pji probability, when reading malicious emails from a adjacent user j. Ni denotes the set of adjacent nodes of node i after eradicate the unknown nodes. Then, we can calculate s (i, t) as in,

$$s(i,t) = \prod_{j\in N_i}[1 - p_{ji}.P(X_j(\tau) = Act)] \quad (12)$$

Where in the event $X_j(\tau) = Act$means that the node j is infected and propels a malicious mail copy to adjacent nodes at time.

Let us consider that the variable τ may take the different values, the equation (12) is disassembled by not including t1 from the range of value τ. There are two cases: First, the user does not checking new emails at time t1. Thus, we have

$$\prod_{j\in N_i}[1 - p_{ji}.P(X_j(\tau) = Act)] = \prod_{j\in N_i,\tau\neq t-1}[1 - p_{ji}.P(X_j(\tau) = Act)] \times \prod_{j\in N_i}[1 - p_{ji}.P(X_j(t-1) = Act)] \quad (13)$$

$$= (1 - s(i,t-1).\prod_{j\in N_i}[1 - p_{ji}.P(X_j(t-1) = Act)] \quad (14)$$

Second, the user checks new emails at time t1. Thus, the malicious email copies are received at time t and those are delivered at time t1by the infected adjacent users. Here we have,

$$\prod_{j\in N_i}[1 - p_{ji}.P(X_j(\tau) = Act)] = \prod_{j\in N_i}[1 - p_{ji}.P(X_j(t-1) = Act)] \quad (15)$$

Actually, the difference of equations (14) and (15) are caused by user checks new emails at time t1. Then unified expression of (14) and (15) is given below:

$$\prod_{j\in N_i}[1 - p_{ji}.P(X_j(\tau) = Act)] = [1 - s(i,t-1).(1 - P\,(open\,_i(t-1) = 1))]\,.\,\prod_{j\in N_i}[1 - p_{ji}.P(X_j(t-1) = Act)] \quad (16)$$

Now, the equation (12) becomes,

$$s(i,t) = 1 - [1 - s(i,t-1).(1 - P\,(open\,_i(t-1) = 1))] \times \prod_{j\in N_i}[1 - p_{ji}.P(X_j(t-1) = Act)]. \quad (17)$$

In equation (17), various trials of $P(X_j(t-1) = Act$) and Ni may guide to different distribution performance.

SIMULATION AND RESULTS

In proposed SEII model the valuation is based on the open analytical model. The spread of the majority email malware is typically impractical to track spreading of malicious mail. In this work, we construct the topology according to the previous investigation of existent email networks. The degree for every node was imitated by the Power-law distribution. The probability of users infected by adjacent nodes (p_{ij}), the email checking time (T_i) and the event generating period (R_i) are determined by human factors. These parameters follow the Gaussian distribution which may provide idealistic values, such as $p_{ij} < 0$ and $T_i < 1$. Here, these values are substituted with their practical range. Thus, if $P_{ij} < 0$, $T_i < 1$ and $R_i < 1$, we let $p_{ij} = 0$ $T_i = 1$ and $R_i = 1$. First, the accuracy for different circulation of p_{ij} is evaluated. The Ti and Ri pursue Gaussian distribution N(40,20^2). As shown in **[Figure- 2]**, the outcomes of SEII model are close to the outcomes of simulations. The SEII model reaches better presentation in correctness.

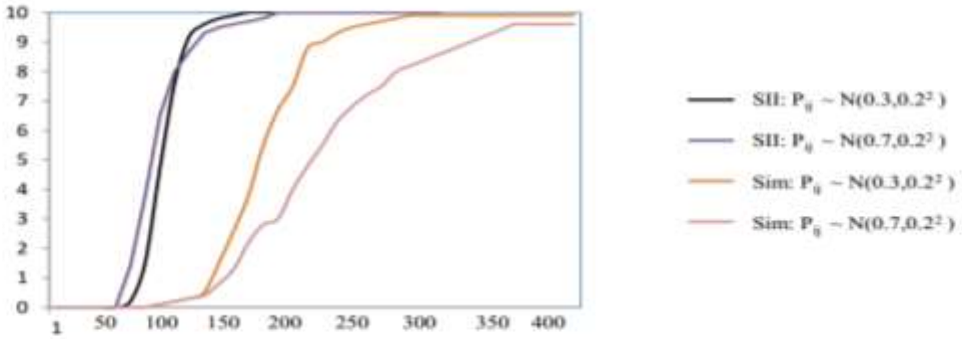

Fig: 2. The accuracy with different distributions

Second, the accuracy for different topologies is evaluated. Ti and Ri pursue Gaussian distribution N(40,20^2) and the probability of infection pij follow N(0.5,0.2^2). As shown in **[Figure- 3]**, the proposed SEII model is efficient in various topologies with various power-law exponents α and means of degrees E(D).

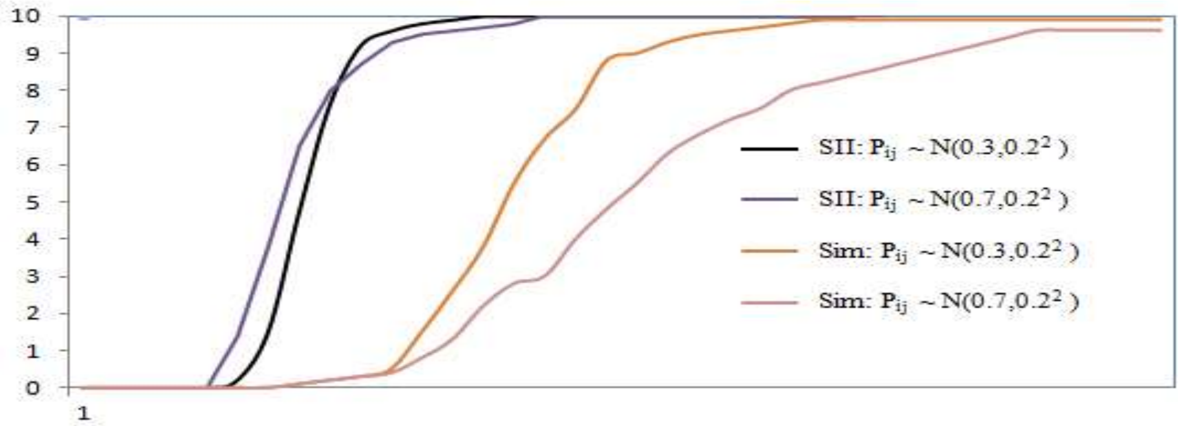

Fig: 3. The accuracy with different distributions of Ti and Ri

CONCLUSION

In this work, we proposed a novel SEII model for evaluate the modern email malware propagation. This model is capable to deal with two critical issues, which are reinfection and self-start. By establishing difference equations, the repetitious distribution processes caused by the above mentioned issues are presented. The outcome of the proposed SEII model is closest to the simulations. For the future work, spatial dependence and temporal dynamics problems are taken into account. Novel simulations have to be designed to enclose real system samples, to analyze the behavior of malware against these samples. The aim of this simulation is to avoid system before being infected by real malware.

CONFLICT OF INTEREST

The authors declare no conflict of interests.

ACKNOWLEDGEMENT

None

FINANCIAL DISCLOSURE

None

ISSN: 0976-3104

REFERENCES

[1] Fan W, Yeung KH. [2011] Online Social Networks-Paradise of Computer Viruses, Physica A: Statistical Mechanics and Its Applications. 390(2):189-197.

[2] Garetto M, Gong W, Towsley D. [2003] Modeling malware spreading dynamics," in Proc. INFOCOM'03. San Francisco, CA. 3:1869–1879.

[3] Wen S, Zhou W, Wang Y, Zhou W, Xiang Y. [2012] Locating Defense Positions for Thwarting the Propagation of Topological Worms, IEEE Comm. Letters. 16(4):560-563.

[4] Xiong J. [2004] Act: Attachment Chain Tracing Scheme for Email Virus Detection and Control, Proc. ACM Workshop Rapid Malcode (WORM '04). 11-22.

[5] Zou CC, Towsley D, Gong W. [2007] Modeling and Simulation Study of the Propagation and Defense of Internet E-Mail Worms, IEEE Trans. Dependable and Secure Computing. 4(2):105-118.

[6] Wen S, Zhou W, Zhang J, Xiang Y, Zhou W, Jia W.[2013] Modeling Propagation Dynamics of Social Network Worms, IEEE Trans. Parallel and Distributed Systems. 24(8):1633-1643.

[7] Calzarossa M, Gelenbe E. [2004] Performance Tools and Applications to Networked Systems: Revised Tutorial Lectures. Springer-Verlag.

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

[9] Sheng W, Yang X, Weijia J. [2014] Modeling and Analysis on the Propagation Dynamics of Modern Email Malware, IEEE transactions on dependable and secure computing. 11:(4)

[10] Sneha S, Malathi L, Saranya R. [2015] A Survey on Malware Propagation Analysis and Prevention Model, International Journal of Computer Applications (0975 – 8887).131(11)

[11] Zou CC, Towsley D, Gong W. [2007] Modeling and Simulation Study of the Propagation and Defense of Internet E-Mail Worms, IEEE Trans. Dependable and Secure Computing. 4(2):105-118.

[12] Gao C, Liu J, Zhong N. [2001] Network Immunization and Virus Propagation in Email Networks: Experimental Evaluation and Analysis, Knowledge and Information Systems. 27:253-279.

[13] Chen Z, Ji C. [2005] Spatial-Temporal Modeling of Malware Propagation in Networks, IEEE Trans. Neural Networks. 16(5):1291-1303.

[14] Wang Y, Chakrabarti D, Wang C, Faloutsos C. [2003] Epidemic Spreading in Real Networks: An Eigenvalue Viewpoint, Proc. 22nd Int'l Symp. Reliable Distributed Systems (SRDS). 25-34.

[15] Yanping Z, Tingting S, Shu Z. [2012] A Novel Model to Restrain Email Virus Propagation, IEEE International Conference on Granular Computing.

[16] Ganesh AJ, Massouli L, Towsley DF. [2005] The Effect of Network Topology on the Spread of Epidemics, Proc. IEEE INFOCOM '05:1455-1466.

[17] Yan G, Eidenbenz S. [2009] Modeling Propagation Dynamics of Bluetooth Worms (Extended Version),"IEEE Trans. Mobile Computing. 8(3):353-368.

[18] Boguna M, Pastor-Satorras R, Vespignani A. [2003] Epidemic Spreading in Complex Networks with Degree Correlations, Lecture Notes in Physics. 625:1-23.

[19] Simarleen K, Arvinder K. [2016] Detection of Malware of Code Clone using String Pattern Back Propagation Neural Network Algorithm. 9(33).