A Survey on Malware Propagation Analysis and Prevention Model

Sneha S.
Student, Department of CSE,
Vivekanandha College of Engineering for Women,
Tiruchengode, Namakkal,
Tamilnadu, India

Malathi L.
Assistant Professor,
Department of CSE,
Vivekanandha College of Engineering for Women,
Tiruchengode, Namakkal,
Tamilnadu, India

Saranya R.
Student, Department of CSE,
Vivekanandha College of Engineering for Women,
Tiruchengode, Namakkal,
Tamilnadu, India

ABSTRACT
In recent years, the security threats imposed by email-based malware, modeling the propagation analysis and prevention of email malware becomes a fundamental technique for predicting its potential damages and developing effective countermeasures. Compared to earlier versions of mail malware, modern email malware exhibits two new features. One is reinfection and another one is self-start. In reinfection, whenever any healthy or infected recipients open the malicious attached file the modern email malware sends its copy to the recipients contact. In self-start, whenever compromised computers restart or malicious files are visited the malware spreads over the system. To avoid these types of issues the security specialists use some of the possible techniques and methods to stop and remove the threats. At the same time the malware developers exploit new malware that bypass implemented security features. In this paper, we analyzed the malware propagation and detecting mechanisms. This survey paper highlights the existing detection and analysis methodologies used for these malicious code.

Keywords
Email malware, SIS Model, SIR Model, SII model, ACT, SEIR Model.

1. INTRODUCTION
Email has become an indispensable communication form in social network. Lacking of knowledge of email virus, a number of users have great confidence on email sent by their friends which results in virus outbreaks in network. The security of Email has a great effect on the security and steady of Internet. Therefore, it's necessary to study virus propagation model. "Emails are widely used to synchronize real-time communication, which is inconsistent with its primary goals". Email messages are designed to be sent, accumulate in repository and be periodically collected and read by receipt. Since most people rely on emails for efficiency and effectiveness of communication, mail boxes may become congested. Messages range from static organization knowledge to conversations with such a broad horizon of messages. Users may find it difficult to prioritize and successfully process the contents of new incoming messages. Also it may be difficult to find a previously archived message in the mail box. At this stage new effective method for managing information in email, reducing email overloads is developed by classifying emails based on importance of words in the email messages and deriving the closest classes the mail could belong to either: critical, urgent, very important, important and not important.

2. PROPAGATION MECHANISMS
A computer virus is one of the major forms of malicious information spreading in the Internet. According to their propagation mechanisms, computer viruses are categorized into scanning-based viruses and topological-based viruses [1].

For scanning-based viruses, Internet users' computers are infected when they have vulnerabilities in their operation systems or installed software. Once a computer is infected, it will send out specific packages to randomly generated targets looking for new victims who have the same vulnerabilities. Once a susceptible computer has been detected, a virus copy will be transferred to this newly detected victim. The efficiency of scanning-based viruses depends on the scanning function of new victims. It has been proven that the preferential scanning mechanism currently has the best efficiency in spreading a virus [2]. The virus using preferential scanning mechanism always considers computers residing in neighboring IP addresses are more likely to be potential victims. A typical instance of this type of virus is Code Red, which successfully infected millions of computers in the Internet [3].

Topological-based viruses become a critical threat to the Internet because online social networks are attracting more and more users in recent years. Once an Internet user is infected by a topological virus, the computer of this user will send malicious email copies to friends embedded in email lists or post infectious hyperlinks on the wall of online social network platforms. When users receive and read the malicious email copies or visit the malicious webpage conducted by the hyperlink, their computers will be infected. The infection processes are repeated from one user to their topologically neighboring users, and then spreads quickly, reaching a large scale. Typical instances of topological-based viruses include the 'Love Letter' email worm and 'KoobFace' that spread in Facebook.

There are four classes of propagation models, defined by whether infected users can become susceptible again after recovery. If this is true, the models are called SIS models because users can change their status as Susceptible-Infected-Susceptible (SIS) [4]. If infected, users cannot become

susceptible again once they are cured, and these models are called Susceptible-Infected-Recovered (SIR) [5,6,7] models. Users can only have a status transition as SIR or Susceptible-Infectious (SI) models [8] if no infected users can recover. If susceptible users can be directly immunized and never become infected, the models are called Susceptible-Infected-Immunized (SII) models [9,10,11,12].

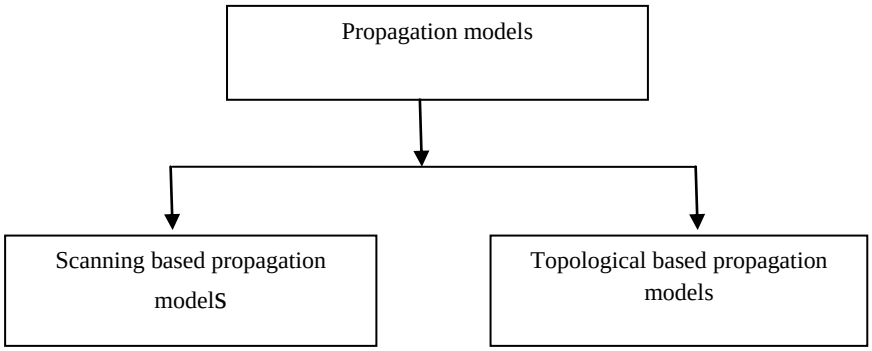

Fig 1: Malware Propagation Models

3. LITERATUR SURVEY

3.1 ATHDI

An Anti-virus and Trust level Healthy-Danger-Infected model was introduced by Yanping Zhang[13]. Many researchers have been studying the characteristics of email virus. Besides, they also have many features, such as hidden, destructive, and so on. Because of the privacy of email, it was impossible to detect whether virus were included in email before receivers receive them. Therefore, antis-virus ability of nodes and trust level between users are particularly important. Such as don't easily open email attachments from strangers (low trust level), and properly open email attachments from familiar friends (high trust level). These measures can greatly reduce the probability of infected. The concept of 'antis-virus ability' and 'trust level' to decrease the number of infected individuals. When a user checking email, users installed anti-virus software will resist virus rather than eliminate all virus. If a user opened virus-embedded email, the node will be infected and send its virus attachments to its neighbors. However, neighbor users can determine to accept or reject the virus email according to trust level between users. This model includes similar characteristics to empirically observed virus propagation processes. The propagation process contains two stages: rapid spread and stable state. It compares the effect of important factors on infected individuals in different networks

3.2 SEIR MODEL

The SEIR Model was introduced by Chao Wang[14] to analyze the virus propagation on famous SNS networks, like Facebook. This model mainly focuses on is undirected network, which will be described by a figure. Each node in the figure stands for a user. The edge connects node i and node j represents that user i and user j are friends. This model defines the number of users on one's friend list as the degree of this node. According to the propagation rules of SNS virus, we divide all of the nodes into four categories: susceptible (S), exposed (E), infectious (I) and recovered (R). Susceptible nodes represent those who are capable of contracting virus; exposed nodes represent those who are infected but not yet infectious; infectious nodes represent those who are infected and capable of transmitting the disease; and recovered nodes represent those who are permanently immune. SEIR Model defines virus propagation rules as follows:

- If a susceptible node contacts with an infectious node, then the probability of the susceptible node transmits into an exposed node is P.
- An exposed node will transmit into an infectious node with the velocity ε without contact with any other nodes.
- An infectious node won't endlessly spread virus.
- An infectious node will transmit into a recovered node with the velocity γ, without contact with any other nodes.

The change of nodes' states while SNS virus spreading is depicted in Fig.2

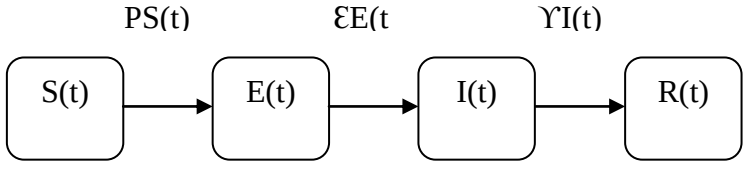

Fig.2 SEIR Model

3.3 ACT SCHEME

An ACT Scheme was introduced by Jintao Xiong[15]. Transmission chain is the tool used to confirm virus spread. In concept of transmission chain, they first define different layers of contact of a host i. In an email network, the hosts that have epidemiological links with host i are the hosts whose email addresses are stored on host i. For example, the neighbors of a host are the primary contacts of the host.

A transmission chain starts with an index case. An index case is a case which shows infectious symptom but has no epidemiological links from any other suspicious or probable case. When an index case is identified, contact tracing is then used to monitor the development of a possible transmission chain from the index case. The transmission chain identification and control system consists of four major processes: the case finding process (CFP), the transmission chain (TC) management process (TCMP), the quarantine process (QP) and the immunization and treatment process (ITP). The diagram of the system is shown in Fig.3. The case finding process is responsible for detecting hosts with infectious symptom. The TC management process is

responsible for setting up, updating, finishing and removing transmission chains. The quarantine process is responsible for controlling the propagation of virus. The immunization and treatment process is the vaccination and virus cleaning process.

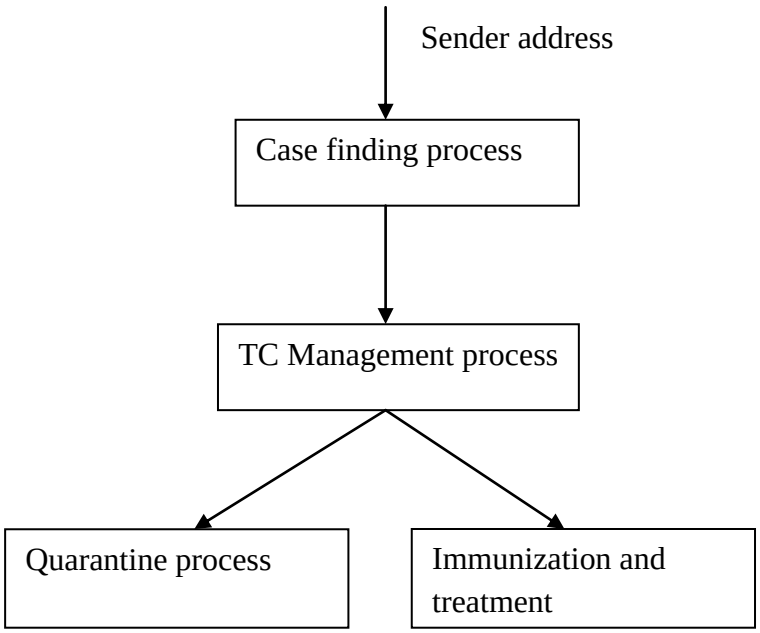

Fig.3 ACT System Diagram

3.4 SIR Model

SIR Model was introduced by B.Rozenberg and this model mainly focuses on the worm propagation on the email and social networks. There are two main strategies in email worms propagation, 1)reinfection 2)non-reinfection. Reinfection strategy defines whenever any healthy or infected recipients open the malicious attached file the modern email malware sends its copy to the recipients contact. But in non-reinfection strategy the infected user sends the worm to his neighbors only once. This SIR model describes the non-reinfection strategy of the email worm. This epidemic spreading model categorize the population into three states:

Susceptible(S) - individuals that are vulnerable and can possibly be infected.

Infected(I) - individuals that have been infected and infect other individuals.

Removed(R) - infected individuals that do not infect other individuals.

For homogeneous networks, dynamics of propagation of email worms can be approximated by SIR model for homogeneous network.

3.5 SII Model

This new analytical model was proposed by Sheng Wen[16] to capture the interactions among the infected email users by a set of difference equations, which together describe the overall propagation of the modern email malware. Then they introduced a new concept of virtual nodes to address the underestimation in previous work, which can represent the situation of a user sending out one more round of malware copies each time this user gets infected.

3.5.1 Virtual Node Generation

The basic elements for the propagation of modern email malware are nodes and topology information. A node in the topology represents a user in the email network. The Virtual nodes, which can represent the situation of a user sending out one more round of malware copies whenever this user gets infected. 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. Malware emails are also sent out when certain events are triggered. Thus, 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, they introduce virtual nodes to present the kth infection caused by infected users opening the kth malware email copy. This model is able to address two critical processes unsolved in previous models: the reinfection and the self-start. By introducing a group of difference equations and virtual nodes, we presented the repetitious spreading processes caused by the reinfection and the self-start.

Nodes and topology information are the basic elements for the propagation of modern email malware. A node in the topology represents a user in the email network. Let random variable Xi(t) denote the state of a node i at discrete time t. Then,

$$X_i(t) = \begin{cases} \text{Hea.}, \text{Healthy} \begin{cases} \text{Sus.}, \text{Susceptible} \\ \text{Imm.}, \text{Immunized} \end{cases} \\ \text{Inf.}, \text{Infected} \begin{cases} \text{Act.}, \text{Active} \\ \text{Dor.}, \text{Dormant} \end{cases} \end{cases}$$

All nodes in networks are initially susceptible. Since infected users will send out malware copies when they are compromised, node i transits from the susceptible state to the active state after the user of node i gets infected. The user is infectious at the active state. When a user is infected but not infectious, the node of this user transits to the dormant state. Besides, any user can be compromised again even if the user

has been infected before. Whatever the state an arbitrary node is at, it may transit to the immunized state.

Table 1. Comparison of various propagation models

S.No	Propagation Model	Result	Advantages	Disadvantages
1.	Anti-virus Trust level Healthy-Danger-Infected (ATHDI)	Improved model that includes anti-virus program and trust- level of user.	More factors are considered for propagation analysis	Impossible to detect the presence of email before the receivers receive the email.
2.	Susceptible-Exposed-Infectious-Recovered	Virus propagation is faster while users login in their accounts more frequently.	Indicate the effects of the networks and users on SNS network.	Assess of the anti-virus strategy is low.
3.	ACT Scheme	Propagation of the virus in the network is controlled by identifying the existence of the transmission chain in the network.	It uses the contact list tracing to find the epidemiological links between host.	Limited to single enterprise, where it is possible to collect all necessary traffic information in a casual chain.
4.	Susceptible-Infected-Removed (SIR)	Dynamic propagation of email virus was approximated for homogeneous network.	Used in email and social network.	Not suitable for heterogeneous network.
5.	Susceptible-Infected-Immunized (SII)	This model is able to address two critical processes such as reinfection and self-start that is unsolved in the previous models.	Avoid systems threats before infected by virus.	Independent assumption and periodic assumption are unsolved problems

4. CONCLUSION

Malware is posing a threat to users computer systems in terms of stealing personal and private information, corrupting or disabling our security systems. This paper highlights some existing methodologies incorporated by security researchers to tackle these threats. Our survey paper explains about different propagation techniques. The malware developer tries to write new techniques and strategies to hide the malicious code and infect the targets. On the other hand, the detectors analyze malware behaviors continuously and try to resist these techniques and strategies hence, we need to allow detection development techniques to lead malware updating through very well analytical process for malware activities and behaviors to fix any possible targeted threats. A new simulation must be designed to contain real system samples, to analyze the malware behaviors against these samples after elaborate malware updating. The objectives of this simulation are to avoid systems threats before being infected by real malware.

5. REFERENCES

[1] Y. Wang, S. Wen, Y. Xiang, and W. Zhou. ‘Modeling the propagation of worms in networks: A survey’, Communications Surveys Tutorials, IEEE, PP(99):1–19, 2013.

[2] M. Vojnovic, V. Gupta, T. Karagiannis, and C. Gkantsidis. ‘Sampling strategies for epidemic-style information dissemination. Networking’, IEEE/ACM Transactions on, 18(4):1013–1025, 2010.

[3] C. C. Zou, W. Gong, and D. Towsley. ‘Code red worm propagation modeling and analysis’, In Proceedings of the 9th ACM conference on Computer and communications security, CCS’02, pages 138–147, New York, NY, USA, 2002.

[4] R. Pastor-Satorras and A. Vespignani. ‘Epidemic spreading in scale-free networks’, PHYS.REV.LETT., 86:3200–3203, 2001.

[5] M. Boguna, R. Pastor-Satorras, and A. Vespignani,’Epidemic spreading in complex networks with degree correlations’, Lecture Notes in Physics, pages 1–23, 2003.

[6] Y. Moreno, J. B. G´omez, and A. F. Pacheco, ‘Epidemic incidence in correlated complex networks’ , Phys. Rev. E, 68, Sep 2003.

[7] Y. Moreno, R. Pastor-Satorras, and A. Vespignani. ‘Epidemic outbreaks in complex heterogeneous networks’, The European Physical Journal B, 26(4):521–529, Apr. 2002.

[8] C. C. Zou, D. Towsley, and W. Gong, ‘Modeling and simulation study of the propagation and defense of internet e-mail worms’, IEEE Transactions on Dependable and Secure Computing, 4(2):105–118,2007.

[9] Z. Chen, L. Gao, and K. Kwiat, 'Modeling the spread of active worms', In INFOCOM 2003. 22th IEEE International Conference on Computer Communications. Proceedings, pages 1890–1900, 2003.

[10] S. Wen, W. Zhou, Y. Wang, W. Zhou, and Y. Xiang,' Locating defense positions for thwarting the propagation of topological worms'. Communications Letters, IEEE, 16(4):560–563, 2012

[11] S. Wen, W. Zhou, J. Zhang, Y. Xiang, W. Zhou, and W. Jia, 'Modeling propagation dynamics of social network worms', Parallel and Distributed Systems,IEEE Transactions on, 24(8):1633–1643, 2013.

[12] C. C. Zou, W. Gong, and D. Towsley, 'Code red worm propagation modeling and analysis'. In Proceedings of the 9th ACM conference on Computer and communications security, CCS'02, pages 138–147, New York, NY, USA, 2002.

[13] Yanping Zhang, Tingting Sun and Shu Zhao, 'A Novel Model to Restrain Email Virus Propagation', IEEE International Conference on Granular Computing, 2012.

[14] Chao Wang, Ke Xu and Gaoyu Zhang, 'A SEIR-based model for virus propagation on SNS', IEEE DOI 10.1109/EIDWT, 2013.

[15] Jintao Xiong, 'ACT: Attachment Chain Tracing Scheme for Email Virus Detection and Control', October 29, 2004.

[16] Sheng Wen, Yang Xiang and Weijia Jia, 'Modeling and Analysis on the Propagation Dynamics of Modern Email Malware', IEEE transactions on dependable and secure computing, vol. 11, no. 4, July/August 2014.

[17] Ms Ranjani.R, Mrs L.Malathi, "SIP Flooding Attack Detection Using Hybrid Detection Algorithm" in "International Journal of Modern Trends in Engineering Research" vol 01, issue 05, Nov 2014.