Analysis of Internet Malware Propagation Models and Mitigation Strategies

Aliyu Mohammed
Universiti Teknologi Malaysia
Faculty of Electrical Engineering

Sulaiman Mohd Nor
Universiti Teknologi Malaysia
Faculty of Electrical Engineering

Muhammad Nadzir Marsono
Universiti Teknologi Malaysia
Faculty of Electrical Engineering

***Abstract*— the Internet application has been so popular in the recent times and the concepts of network worms are now the biggest threats to the network security researchers. The trend at which network structure variables and its related applications such as p2p network has given the malware on the internet an avenue to explore as rapidly as possible and to find ways from which they can propagate. Hence, the need for an accurate mathematical worm propagation model is of paramount desire for the internet. In this paper a conceptualized analysis of the worm propagation models are discussed for the different prevailing environments. The paper tries to highlight the pros and cons of the models and compares them for possible control strategies that could be achieved. It is the desire that others in the research community requiring to learn more on the speed and the dynamics of the changing worm literature can find it interesting to look at.**

Keywords- Internet worm; Propagation Model; simulation, Defense Mechanism Introduction

I. Introduction

The internet application has been so popular in the recent times and the network worms are now the biggest threat to the network security. The worm that appeared in the wild was the Morris Worm of 1988 for the first time. While the Code Red and Nimda worms of 2001 infected hundreds of thousands of computers on the network, that brought about some huge amounts of dollars lost to the entire society on the Internet [1]. In 2008, a conficker was detected and it was a computer worm which is designed to target Microsoft Window operating systems. It infected about 15 million hosts and lead to the sinking of French Navy network [2] and it is believed to be largest computer worm infection since outbreak of SQL Slammer of 2003 [3]. The heart of Internet worm system has six component characteristics that either individually or collectively contained in the worm system. These components are categorized as:- *reconnaissance capabilities; specific attacks capabilities; command interface ; communication capabilities; intelligence capabilities and unused attack capabilities* [4]. Researchers develop different models to represent the typical worm spread characteristics. The models are typically categorized into three groups namely: analytical models, simulation models, and hybrid models. Therefore, epidemiologically, these models are base on both stochastic and deterministic models for effective modeling of the spread of infectious disease [5].

II. Classical Epidemic Models

A. Kermack – Mckendrick Model

Based on epidemiological modeling, Kermack Mckendrick (KM) model is referred to as the SIR model of epidemics [6]. It tries to reflect the nodes that are likely to be vulnerable and will be infected by worms as susceptible, while the nodes that have been infected and also have the tendency of infecting others as infectious nodes and the nodes that have gained immunity or dead and don't have the likely hood of being infected again by the worms are termed as removed nodes. The entire population of nodes that are susceptible, infectious, and removed at time **t** is given as **S** (t), **I** (t), and **R** (t) respectively. The KM related models is as shown below:

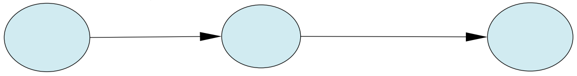

Fig.2.1 A Typical SIR Model

$$\frac{dS}{dt} = -\beta IS \quad ; \frac{dI}{dt} = \beta IS - I\alpha \quad ; \quad \frac{dR}{dt} = \alpha I$$

Where **β** and **α** are the infection rate and the removal rate. The entire population is **N,** and the ratio S (t) + I (t) + R (t) = 1.*A typical propagation model for the SIR equation above is as indicated in the figure 2.1 adapted from D. Smith and L. Moore* [8].

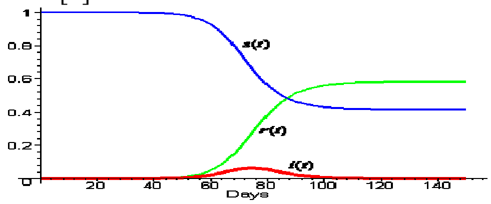

Fig. 2.1 Basic SIR Propagation Model

Generally, the model considers the removal rate process as the factor that relates to the action which involves human countermeasures, at the same time it has its drawbacks which includes the possibility of network congestion and topological changes. *Thus with large – scale worm propagation it can cause congestion problem and tends to trouble the routers,*

eventually slows down the worm scanning process. With variable infection rate and dynamic hosts could address the drawbacks of the model as handled by the Two-Factor model.
”.

B. The Two – Factor Model

The previous worm propagation models [9], tend to neglect the possible dynamic effects on the network which include the human countermeasures on the worm behavior and the changing magnitude of the infection rate on the network.

The two-factor [10] model treated the effects of the SIR based on the following parameters, thus the name two-factor model:-

Human countermeasures, leads to removing both the susceptible and infectious hosts from the network circulation as the case of Code Red worm propagation. As the hosts becoming aware of the worm propagation and the implemented countermeasures like cleaning compromised hosts, patching or upgrading susceptible hosts, setting up filters to block the worm traffic on firewalls or edge routers or some time disconnecting the hosts from the internet [5].

Decreased infection rate β: infection rate β is not a constant parameter, as in the case of the ideal internet which is dynamic. Thus with large – scale worm propagation it can cause congestion problem and tends to trouble the routers, eventually slows down the worm scanning process [5].

In effect, the complete two-factor worm propagation model equation is typically of the form indicated below:-

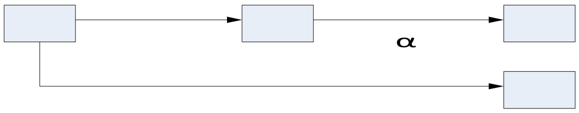

Fig. 2.2 A Typical Two- Factor Model

$$\frac{dS(t)}{dt} = -\beta(t)S(t)I(t) - \frac{dQ(t)}{dt} \quad ; \quad \frac{dR(t)}{dt} = \alpha I(t) \quad ;$$

$$\frac{dQ(t)}{dt} = \mu S(t)J(t)$$

$$\beta(t) = \beta_o\left[1 - \frac{I(t)}{N}\right]^{\eta} ; \quad N = S(t) + I(t) + R(t) + Q(t)$$

Where $J(t)$ the number of infectious hosts including removed hosts at time t, and $Q(t)$ is the number of removed hosts from susceptible hosts at time t. where exponent η is used to adjust the infection rate sensitivity to the number of infectious hosts I (t)[10].

The two-factor model can be employed to analyze some large – scale worm propagation systems but it has its drawback of not having effect of topological influence on the worm propagation. This is for the fact that it specifically considers Code Red mode of propagation on the network environment.

III. The Development of Internet Worm Propagation Models.

The journey so far in the development of Mathematical worm propagation modeling is chronologically articulated in the Table1 below:

Table1. Worm Propagation Models.

ARTICLE AUTHOR	METHODS/ STRENGTH	WEAKNESS
Kermack & MC 1993	**SIR; Epidemic model**	No birth/death population .
C.C.Zou et.al 2002	**Two – Factor Model; no topology Constrains, Code Red worm**	Modeling only continuously spreading worms, no stop and start
Wang &Wang 2003	**SIS /SIR based on K.W model**	Only effective for SIS model
Zou et al 2003	**SIR model with Dynamic Quarantine; active worm**	Not suitable for non-active worms
Chen et al-2003	**AAWP- random scanning worm propagation**	Did not consider topological- scanning worms
Zhang et al-2004	**Based on SIR model, reduce propagation capabilities**	Infection delay and topology
Yuan,H, et al 2008	**e-SEIR-extended, impact of quarantine ,latency**	Epidemiological modeling based
Zhao,N. & Zhang,X 2008	**SIHR- based on Honeynet -based on Two –Factor, immunity control**	Known topology, Honeynet spreading worm only.
ARTICLE AUTHOR	METHODS/ STRENGTH	WEAKNESS
Jin,C et al 2009	**Improved SEIR model, vigilance and removing time control**	Best suit email virus propagation mostly.
Fangwei Wang, et al	**SEIQV- model, reduced speed,**	Could not account for the dynamic topological

2009	**infection**	change
Chen ,J ; Shengjun W 2009	**SIRS- model , based on IP**	Could not account for strong immunity of removal.
Fei, SU et al 2009	**Two-Factor Model as ETwo- Factor, effects of NAT ,Firewalls**	Difficulties of handling complicated networks.
Wang, Yuanmei et al. 2010	**Improved SIR with dynamic i/0 property**	Only for homogeneous networks with impulsive effects.
Li, Tao et al 2010	**SIR epidemic model , non-linear incidence rates**	Controlling the spread of non-linear incidence rates of the system network
Wang, Juan et al 2010	**SEIRQ model, immunity**	Could not consider the dynamics of the topology
Yao ,Yu et al 2010	**SIQV model ,analytical &simulation**	The strategy doesn't include topological dynamics.
Fangwei , W. et al 2010	**SEIQV model, for Slammer worm**	Limited Topological dynamic changes
D. Zhang & Ye Wang 2010	**SIRS model, no permanent immunity**	No permanent immunity and linear dynamics of propagating worms

A. *The SIR model: –*

An improved Kermack –Mckendrick model with dynamic input – output and time delay [11] to handle the issues of nodes or hosts that are infected and cured becomes permanently immunized. Thus the model considers the dynamic of the hosts to be under temporary immunity, as the nodes can be infected again with the same worm or another. There was a concise stability of the worm propagation model based on homogeneous network environment. In [12], the propagation model considers the dynamic input/output property within a non linear incidence rate analysis. It provides for an impulsive control mechanism on the spread of the worms. In [13] the impulsive effect on the dynamic input/output to establish an asymptotic stability of the worm propagation on the homogeneous network. The bilinear incidence rate [14] was considered to improve the classical Kermack – Mckendrick (KM) model, where a necessary condition for threshold to the existence of equilibriums were established. The model proffers for further study into the dynamic increase and decrease rate of hosts and quarantine strategy in the network. In [15],the worm propagation time delay due to the use of anti – virus software was able to improve the model which was based on the quarantine and human countermeasure, similar to the two – factor model.

B. *The SIRS Model: -*

Based on the KM model, the work in [16] handles the issues of the dynamics of the network. The dynamics of the network provides for increased removal rate and other hosts that are partially removed and could become susceptible again to be infected. The increased removal rate provides for better understanding and gave effective measures to use for controlling the spread of the internet worms. In [17], the model was based on the fact that some hosts on the network can be partially immunized and can still be infected by worms within or outside the environment. This is as a result that they only updated anti-virus without patching and loophole plugging.

C. *The Two – Factor Model: -*

The improved two factor model of [5] was derived on the context of ETwo-Factor model that is charged with the ability to handle worm propagations like the Code Red II. The model is based on non-homogeneous systems. It indicates that interactions across groups of hosts will not be the same with interaction within groups, thus it provides for worm propagation on two different network topologies. It also indicates that locality propagation of worms is going to resemble the situation where most vulnerable hosts tend to be placed across the entire network. Eventually, the system could not pair fairly well with some complicated network topologies.

It could be recalled that the two-factor model considers all hosts on the internet to be identical, which is not an accurate assumption; this is due to the fact that there exist on the network the effect of network address translation (NAT) [18]. Network Address Translation is a mechanism used in replacing IP address information in the packet headers while in transit across a traffic routing device for the purpose of trying to remap the given address space available in another IP. The model TLWM (three layer worm model) is based on the NAT which provide the environment to represent the worm propagation. The three layers comes as a result of the hosts and routers on the internet, the given NAT hosts, and the hosts under the NAT distribution as layers 1, layers 2 and layers 3, respectively. In [19], the improvement of the SIR model addresses the effect of new host state called exposed hosts and the variable infection rate, thus it is termed ETwo-Factor. The model considers the commonly ignored issues like temporary immunity and latency of worms which could lead to greater form of destruction.

In summary, the worm propagation models are basically into the three categories; SIR, SIRS, and Two-Factor. The SIR has the effect of not considering the birth and death rate as well as the spatial distribution of the vulnerable hosts. The model has been modified by the introduction of dynamic quarantine, vaccination and removal/ infection rates. On the other hand, SIRS is specific to model the non- permanent effects of immunity on the removed hosts. In the Two-Factor model it considers the effects of internet Code Red worm propagation scenario.

IV. Worm Propagation Mitigation Strategies

The idea of modeling worm propagation is to study the effectiveness of the available methods of worm detection and containment. This will enable for finding the right methods to apply and permit for combating the worm's behavior effectively [20]. The model addresses worm containment based on dynamic quarantine on the hosts in a group that have exhibited suspicious behavior. Thus by vaccinating the hosts in the prevailing group, they become immunized to the worm infection and that is the implied model called SEIQV (Susceptible, Exposed, Infectious, Quarantined, and Vaccinated) and provides an enabling guideline for typical active worm control.

In [19], the analysis was on the different containment measures that could be attained through varying the dynamics of the model parameters. On the other hand, [18] provides for the critical look at the NAT environment as it could provide for the development of effective defense techniques that will be based on the worm propagation behavior. Verification of the model parameters could be analyzed through the use of simulation strategy based on the randomness of the propagating malware on the network [21]. The mitigation and control strategies of propagating malware on the network environment have multidimensional factors to enable for understanding their effectiveness.

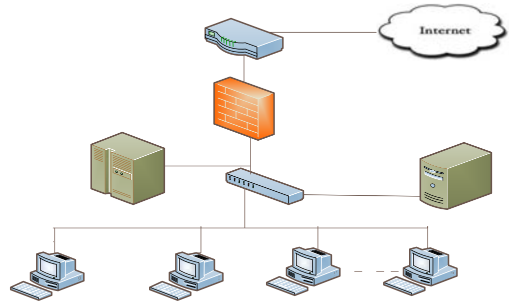

Fig.4.0 Simple Defense Mechanism.

The effect of factors like the use of anti-virus, IPS/IDS, and patch management could provide for a realistic mitigation and control strategies on the network environment. These factors could include the possibility of studying the different defense system strategies in order to provide for evaluation of the various trends of the malware propagation. In [22], a systematic analysis of the complete worm defense strategies were studied and a taxonomy of the key factors each defense category was provided. The rapid spreading nature of Internet worms on the network calls for an automatic mitigation to be implemented. In [23], it provides for a model which combines the dynamic quarantine method with vaccinations. The model tends to improve the population of infectious hosts and reduces the entire worm spreading characteristics. On the other hand, [24] considers the effect of constant quarantine strategy as not quite suitable for typical fast spreading worms; this is as a result of insufficient timing. Thus the model introduces the issue of applying repeated quarantine impulsively for a period of time.

.

Conclusion.

In the aforementioned survey analysis of worm propagation model, it requires that a lot needs to be done in understanding the true worm behavior in spreading over the internet. All the models for the propagation of worm spread have in common same constraint which relates the network topology and/or bandwidth dynamics. The recent models still have some drawbacks that are still not fully addressed, these includes the non-linearity in the dynamics especially true for mobile/wireless nodes(notebook), while NAT tends to provide the environment for enhanced defense mechanisms.

The simulation consequence suffers from the real desire to have basic working model which does not entail that it should be a closed system. But in a situation where a mixed-mode simulation is employed it does provide an effective tool that will enable for the understanding of the worms and virus propagation. The system is such that particular subnets are simulated in details, while the internet as an entire integral is simulated from its aggregated behavior. This is referred to as the environment where the viruses are eminent and propagating at rapid rate.

References

[1] [1] D. Moore, *et al.*, "Code Red : A case study on the spread and victims of an Internet worm," 2002.

[2] [2] M. R, "Conficker worm sinks French navy network," 2008.

[3] [3] S. Staniford, *et al.*, "How to own the Internet in Your SpareTime," *Usenix Security Symposium, San Francisco* 2002.

[4] [4] S. Fei, *et al.*, "A survey of Internet Worm Propagation Models," *IEEE,* 2009.

[5] [5] S. Fei, *et al.*, "Worm Propagation based on Two-Factor Model," *IEEE,* 2009.

[6] [6] D. Smith and L. Moore, "The SIR model for the Spread of Diseases " *MathDC : Dec, 2001,MM,* 2008.

[7] [7] K. jonghyun, *et al.*, "Measurement and Analysisof worm propagation on Internet network topology," *Proceeding of the International Conference on Computer and Communication Engineering,* 2004.

[8] [8] D. Smith and L. Moore, "The SIR model for spread of disease " *Journal of online Mathematics and its applications,* 2001.

[9] [9] C.C.Zou, *et al.*, "On the Performance of Internet Worm Scanning Strategies," *9th. ACM Symposium on Computer and communication Security,* 2002.

[10] [10] C. C. Zou, *et al.*, "Code Red Worm Propagation Modeling and Analysis," *ACM,* 2002.

[11] [11] T. Li, *et al.*, "The Stability of a Worm Propagation Model with Time Delay on Homogeneous Network," *IEEE,* 2010.

[12] [12] T. Li, *et al.*, "Impulsive Control of the Spread of worm with Nonlinear Incidence Rates," *IEEE,* 2010.

[13] [13] Y. Wang, *et al.*, "Modeling and Analyzing the Spread of Worm with Impulsive Effect on Homogeneous Network," *IEEE,* 2010.

[14] [14] C. Junhua and W. Shengjun, "Modeling and Analyzing the Spread of worms with Bilinear Incedence Rate," *IEEE,* 2009.

[15] [15] W. Shaojie, *et al.*, "Analysis of a Mathematical Model for Worm Virus Propagation with time delay," *IEEE,* 2009.

[16] [16] D. Zhang and Y. Wang, "SIRS: Internet Worm Propagation and Application," *IEEE,* pp. 3029-3032, 2010.

[17] [17] Q. Liu, *et al.*, "Modeling and Analysis of an SIRS Model for worm Propagation," *IEEE,* 2009.

[18] [18] S. Fei, *et al.*, "Modeling and Analysis of Internet worm propagation," *Scince Direct,* vol. 17(4), pp. 63-68, 2010.

[19] [19] J. Wang, *et al.*, "A novel Model for the Internet Worm Propagation," *IEEE,* 2010.

[20] [20] F. Wang, *et al.*, "Epidemic models applied to worms on internet," *IEEE,* 2009.

[21] [21] Z. Wei, *et al.*, "The Study of Network Worm Propagation Simulation," *IEEE,* vol. 9, pp. 295-299, 2010.

[22] [22] Liuqi and G. Ma, "The research and development of worm defense strategies," *IEEE,* pp. 168-171, 2010.

[23] [23] F. Wang, *et al.*, "Stability analysis of a SEIQV epidemic for rapid spreading worms," *Computer and Security, ELSEVIER,* pp. 410 - 418, 2009.

[24] [24] Y. Yao, et al., "The Worm Propagation Model with pulse Quarantine Strategy " 2010 Internation Conference on Multimedia Information Networking and Security, IEEE, pp. 269-273, 2010

AUTHORS PROFILE

Aliyu Mohammed is a post graduate student of the university of Teknologi Malaysia, in the Department of Microelectronics and Computer Engineering, in the faculty of Electrical Engineering. His PhD research interest is in networking with reference to network securities.