Proliferation of Worm Infection in P2p Networks and Its Prohibition

B.Manasa Devi
Department of Computer Science & Engineering, Ganapathi Engineering College, Warangal, A.P, India.

Abstract
Malicious software or Malware is the software developed with malicious intentions. Hackers use it for spoiling computer programs or to get access to sensitive information. The detection of such malware can be done by writing program which can understand the dynamics of malware. Towards this end this paper presents an analytical model which can effectively characterize the true nature of malware and how it spreads in P2P networks such as Gnutella. The proposed model is compartmental model which involves derivation of network conditions and system parameters in such a way that under those parameters and conditions the underlying P2P network reaches a malware free equilibrium. The proposed model can also perform evaluation of strategies such as quarantine used to control malware spread. Afterwards the model has been enhanced and tested with networks of smart cell phones. The empirical results revealed that the proposed model is effective and useful.

Index Terms
Malware, peer-to-peer networks, compartmental model, Bit Torrent and Time to Live (TTL).

1. INTRODUCTION

Peer-to-peer networks are networks where there is no specific designation of nodes in the networks. In case of domain network, it is required to designate something as server constantly and other nodes as clients. The P2P model is different from domain model. Peer means a node with same designation. It does mean that in P2P network there is no concept of naming server and client. All the nodes are given equal importance and that is the reason they are known as peers. The usage of P2P networks has become popular and now the usage is spread to various domains which network is possible with certain flexibility. The kind of network required by such systems is the network that has flexible nodes and they are having no much dependency among them. The use of this kind of network has resulted in the flexibility of network connections or services. The proposed network also resembles Gnutella [1] where flooding is the search process. The search process in the network starts with flooding. In this process, a peer forwards the query to its neighbors and then this is repeated until all possibilities are tried in the TTL limits. In the relevant example, the Gnutella systems were affected by the Mandragore Worm. When a node is compromised, it is possible that the compromised node can spread the malware into other systems easier than that of the same without compromised node. The whole search process begins with query in P2P networks. When a node gives query to other node, it is given to a node and gets forwarded if the node is in the TTL limits. The model presented in this paper considers flooding approach as used in Gnutella networks.

2. PRIOR WORK

Many researchers investigated time for achieving the measurement of P2P systems. Such measurement oriented works such as [2], [3] and [4] have good analytical models. This is meant for temporal evaluation of the information available in the network. These works focus on the regular file transformation. However, they are not applied to the malware spreads in the system rapidly. All of them are specialized in networks like Bit Torrent and take time to extend the networks Gnutella, KaZaA and so on. The research papers [5] and [6] address the issue of worms in P2P networks with the help of a simulation study with respect to worms and their neighbor possible migration mechanisms. On the malware study there are some epidemiological models in order to understand the dynamics of spread of malware in decentralized networks of P2P model. However, this assumption looks wrong as it is really invalid. The fact is that the chances of infecting a peer are limited to its TTL hops away from it and not this entire network.

In the models specified, another important behavior not considered is the incorporation of user behavior. In P2P networks, every user has two states. They are known as “on state” and “off state”. As the name implies, the on state indicates that the peers are actively connected to network while the off state indicates that the peers are not actively connected to network. The infecting probabilities are more when peers are in on state. The peers that go offline in the P2P network having less probability of getting infected. In [7] Bit Torrent is considered as an empirical model for malware spreading while the results of infected nodes in the models where dynamic hit list-based malware is considered in Bit Torrent networks and they are shown in [8] and [9]. These models are having a known drawback. It is that they are ignoring dynamics of

Manuscript received September 5, 2014
Manuscript revised September 20, 2014

node like transitions offline and other such models in the real world and the models are applicable to only Bit Torrent networks.

In the researches [10] and [11], the authors used a graph model for P2P networks. From this they derive a limiting condition which limits the scope of the adjacency graph for both virus and worm that is prevalent in the network. The models ignored an important fact that if any node in the P2P network is infected that node is likely candidate to spread malware in the network. However, the compromised node can't infect all the other nodes in the network. It has to be remembered that the infected node can infect any other node which is within TTL hop ratios in the network.

In this paper, we present a good model that demonstrates the dynamic of malware in the P2P network which is modeled after Gnutella. This is capable of overcoming the drawbacks of other networks. As the model has two phases it is possible. In the first phase, the peers in the TTL range are quantified. In the second phase, neighborhood information is taken and considered in studying the final dynamics of malware spread.

3. MALWARE PROPAGATION MODEL FOR P2PNETWORKS

In this paper, this section is an important section as it describes the aspects of the proposed propagation model for Gnutella like P2P networks. The model is assured to ignore regular files and the malware is not ignored. The block diagram of the proposed model is as shown in fig. 1.

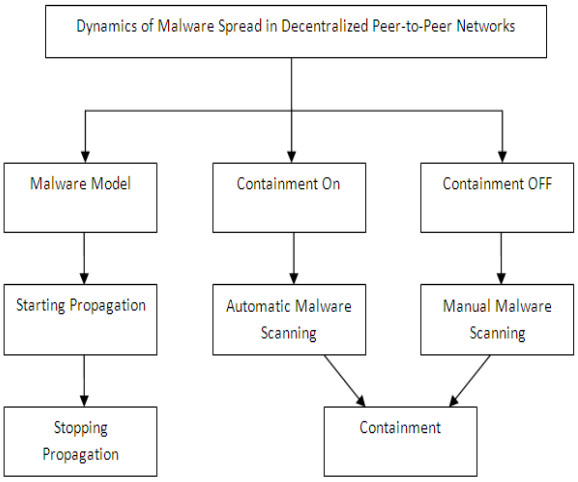

Fig. 1: Block diagram for Proposed Model

3.1 Search Mechanism

In P2P networks, the sharing of information among the nodes is possible first of all by sending the search request. How this request is transmitted over network is important. As per this paper, when a query has to be made, the node is supposed to give the answer. The model used in the network is usage of TTL in the query processing. The query involved TTL bounds and passing message in the TTL bounds. There are two approaches for searching in P2P networks. The first approach is known as flooding where every node sends query to its entire neighbor. When a peer receives affirmative message, it will forward it to next node in TTL. The response of a peer is affirmative if the TTL of the query is greater than zero and then it forwards the query to its neighbors otherwise the query gets discarded. In the proposed system it is assumed that it is sufficient when a file is distinguished as a genuine file or any malware that is sufficient for the proposed system. An approach given in [12] is used now in order to qualify the neighborhood of the search. Then we define the same generating function for PMF (Probability Mass Function).

3.2 Notation and Parameters of P2P Network

The model parameters of the P2P considered for experimentation and their notations are described in table 1.

$\lambda_{on}, \lambda_{off}$	Rate at which off and on peers switch on and off
λ	Rate at which a peer generates queries
$1/\mu$	Average download time for a particular file
r_1	Rate at which peers terminate ongoing downloads
r_2	Rate at which peers renew interest in downloading a file after having deleted it
$1/\delta$	Average time for which a peer stores a file

Table 1: Model parameters of P2P network

3.3 Compartmental Model

The proposed model is made as compartmental model where the peers are classified into compartments. And each node in the network belongs to specific compartment. The system is based on power law topologies. The compartmental model is based on the concept of node degree [13]. The four classes in the compartmental model are described below in table 2.

$P^{(k)}_S$	Number of peers wishing to download a file.
$P^{(k)}_E$	Number of peers, currently downloading the malware
$P^{(k)}_I$	Number of peers with a copy of the

	malware.
$P^{(k)}_R$	Number of peers who either have deleted the malware or are no longer interested downloading any file.

Table 2: Partitioned classes in compartmental model

3.4 Assumptions Made

The following assumptions are made in the proposed system based on the mean-field approach. The assumptions are significant in achieving and characterizing the spread of malware in the decentralized P2P networks and presented here.

- Differential function of time is the number of members in a compartment. It is true for small and big size members present in the compartment.
- The more emphasis is kept on the number of members in each class though differential equations are used in the compartmental model.
- The spread of files in the network is predefined and deterministic. For instance that communication among nodes is starting with search operation and that is based on the flooding approach in the network.
- The size of network is fixed for certain time while making experiments. This is required to characterize the dynamics and spread of malware in the decentralized P2P networks.

With degree "k", the dynamics of malware in the with respect to classes in the compartmental model can be represented by using the following equations.

$$\frac{dP^{(k)}_{Son}}{dt} = \lambda P^{(k)}_{Son}\left(1-(1-p_{inf})^{z(k)}_{av}\right) - r_1 P^{(k)}_{Eon} + r_2 P^{(k)}_{Ron} - \lambda_{off} P^{(k)}_{Son} + \lambda_{on} P^{(k)}_{Soff}$$

$$\frac{dP^{(k)}_{Eon}}{dt} = \lambda P^{(k)}_{Son}\left(1-(1-p_{inf})^{z(k)}_{av}\right) - r_1 P^{(k)}_{Eon} - \mu P^{(k)}_{Eon} - \lambda_{off} P^{(k)}_{Eon} + \lambda_{on} P^{(k)}_{Eoff}$$

$$\frac{dP^{(k)}_{Ion}}{dt} = \mu P^{(k)}_{Eon} - \delta P^{(k)}_{Ion} - \lambda_{off} P^{(k)}_{Ion} + \lambda_{on} P^{(k)}_{Ioff}$$

$$\frac{dP^{(k)}_{Ron}}{dt} = \delta P^{(k)}_{Ion} - r_2 P^{(k)}_{Ron} - \lambda_{off} P^{(k)}_{Ron} + \lambda_{on} P^{(k)}_{Roff}$$

$$\frac{dP^{(k)}_{Soff}}{dt} = \lambda_{off} P^{(k)}_{Son} - \lambda_{on} P^{(k)}_{Soff}$$

3.5 Malware Free Equilibrium

R0 represents the basic reproduction model which is used as a metric that is used to govern the stability of the malware free equilibrium globally. The R0 quantifies the number of vulnerable peers where their security has been compromised by some of the infected hosts in their lifetime. It is very clear in experimental results consideration that if R0 is less than 1, it ensures that the epidemic dies out fast and can't take the endemic state [14]. Stability information is considered very important as it can give guarantee that the system is always malware free even if the newly infected peers are introduced. The following formula is used to achieve MFE in the decentralized P2P network.

$$F = [\partial F_i(x_0)],\ V = [\partial \upsilon_i(x_0)]\ ,\ 1 \le i,j \le m,$$
$$\partial x_j \qquad \partial x_j$$

3.6 Quarantine

Quarantine is nothing but removing infected nodes from the network. By doing so in nodes, it is possible that the limiting the malware spread is done and that is the reason the guaranteed nodes limit the spread of malware. The quantization of quarantine rate is done in this section. The basic reproduction is represented as R0. Quarantine does mean that the node is taken out of network. It is assumed that when nodes are removed from network, they P2P network remains good and does not result into disconnected components.

The following equations represent additional terms.

$$\frac{dP^{(k)}_{Ion}}{dt} = \mu P^{(k)}_{Eon} - \delta P^{(k)}_{Ion} - \lambda_{off} P^{(k)}_{Ion} + \lambda_{on} P^{(k)}_{Ioff} - \eta P^{(k)}_{Ion}$$

$$\frac{dP^{(k)}_{Ron}}{dt} = \delta P^{(k)}_{Ion} - r_2 P^{(k)}_{Ron} - \lambda_{off} P^{(k)}_{Ron} + \lambda_{on} P^{(k)}_{Roff} + \upsilon P^{(k)}_{Q}$$

and the dynamics of $P^{(k)}_Q$ are described by

$$\frac{dP^{(k)}_{Q}}{dt} = \eta P_I{}^{(k)}_{on} - \upsilon P^{(k)}_{Q.}$$

4. RESULTS

This section is used to validate our system through simulation results. The purpose of simulations done is to observe the dynamics of malware spread in decentralized peer-to-peer networks. To achieve this custom simulator is built. The simulation results are analyzed. For thousands of nodes results were simulated and the topology used in power-law topology. As per the system parameters and analytical model described in prior sections, the simulation is carried out and the results were analyzed. Each experiment is performed 20 times and the results are averaged.

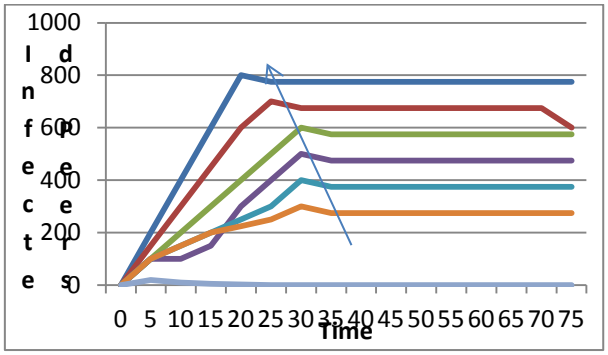

Fig.2: Effect of λonon malware intensity

Fig. 2 shows the results which visualize the time and infected peers. When time grows, the benefits of offline users also more.

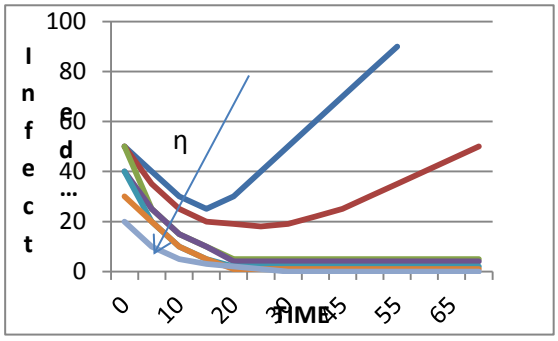

Fig. 3: Effect of quarantine on malware intensity

As can be seen in fig. 3, the effect of quarantine has been plotted. The peers infected is taken in X axis while the time taken for quarantine is given in Y axis.

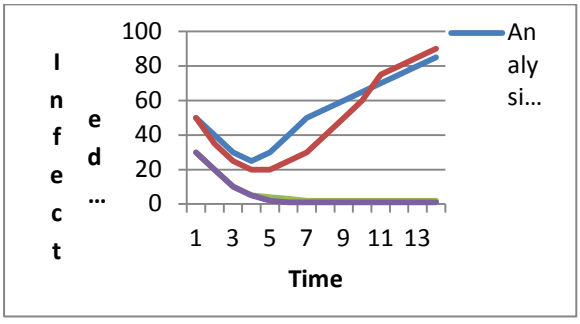

Fig. 4: Effect of quarantine on the system in (29-31) for λ=0.02.

As can be seen in fig. 4analysis and simulation are shown. The infected peers and the time taken for performing quarantine are visualized in fig. 4.

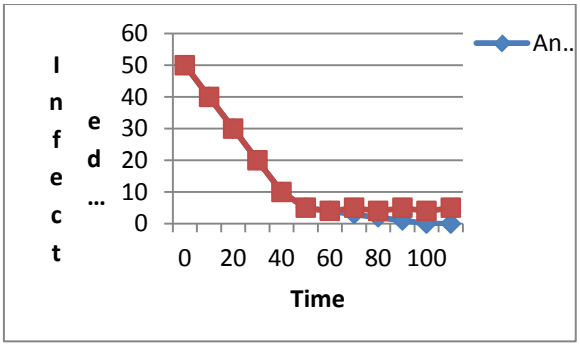

Fig. 5: Impact of λ on malware intensity (λ=0.005) (5-12)

As can be seen in fig. 5, it is evident that it uses the basic reproduction number to be greater than 1. This is assumed to prevail for an epidemic. When R0 is less than 1, the number of infected peers is dropping down to zero.

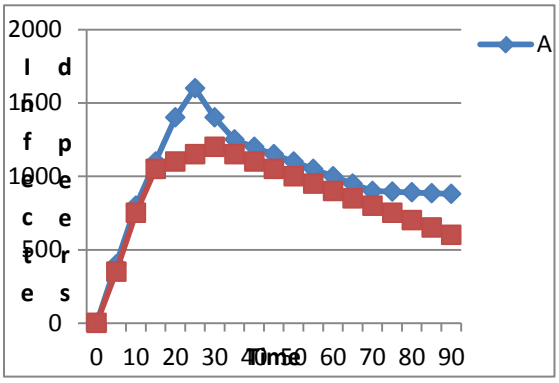

Fig. 6: Impact of λ on malware intensity (λ=2.0) (5-12)

As can be seen in fig. 6, it is evident that it uses the basic reproduction number to be greater than 1. This is assumed to prevail for an epidemic. When R0 is not less than 1, then it reaches epidemic proportions. Malware presences in the nodes that run most of the time online are likely to get infected more when compared the same with offline.

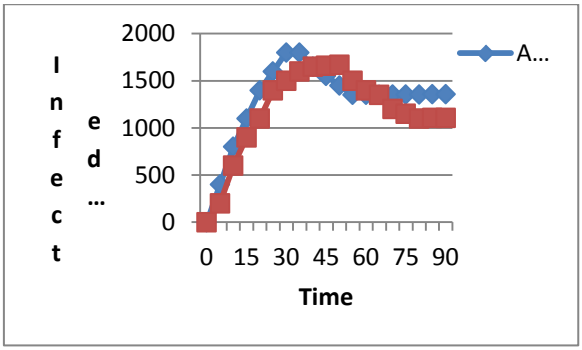

Fig. 7: Influence of offline duration on malware intensity for the system in (5-12)

As can be seen in fig. 7, the peer's infected and time is plotted in Y and X axes. The influence of offline duration on malware intensity could be found.

5. CONCLUSION

In this paper, a model is developed to analyze the spread of malware in Peer – to – Peer networks. The characteristics of malware spreads and its dynamics are incorporated in the analytical model. The model features both offline and online transitional behavior of malware and its dynamics. The proposed model also takes communication patterns for experiments such as size of neighborhood into account. The proposed system also tries to quantify the influence of malware and their ratio in the production. In the estimating of RO the above features can help in accurately modeling the spread of malware. The experiments reveal that the proposed analytical model with certain parameters is capable of proving the efficiency of our model.

REFERENCES

[1] Clip2, "The Gnutella Protocol Specification v0.4," http://www.stanford.edu/class/cs244b/gnutella_protocol_0.4.pdf, Mar. 2001.
[2] X. Yang and G. de Veciana, "Service Capacity in Peer-to-Peer Networks," Proc. IEEE INFOCOM '04, pp. 1-11, Mar. 2004.
[3] D. Qiu and R. Srikant, "Modeling and Performance Analysis of BitTorrent-Like Peer-to-Peer Networks," Proc. ACM SIGCOMM, Aug. 2004.
[4] J. Mundinger, R. Weber, and G.Weiss, "Optimal Scheduling of Peer-to-Peer File Dissemination," J. Scheduling, vol. 11, pp. 105-120, 2007.
[5] A. Bose and K. Shin, "On Capturing Malware Dynamics in Mobile Power-Law Networks," Proc. ACM Int'l Conf. Security and Privacy in Comm. Networks (SecureComm), pp. 1-10, Sept. 2008.
[6] L. Zhou, L. Zhang, F. McSherry, N. Immorlica, M. Costa, and S. Chien, "A First Look at Peer-to-Peer Worms: Threats and Defenses," Int'l Workshop Peer-To-Peer Systems, Feb. 2005.
[7] J. Schafer and K. Malinka, "Security in Peer-to-Peer Networks: Empiric Model of File Diffusion in BitTorrent," Proc. IEEE Int'l Conf. Internet Monitoring and Protection (ICIMP '09), pp. 39-44, May 2009.
[8] J. Luo, B. Xiao, G. Liu, Q. Xiao, and S. Zhou, "Modeling and Analysis of Self-Stopping BT Worms Using Dynamic Hit List in P2P Networks," Proc. IEEE Int'l Symp.Parallel and Distributed Processing (IPDPS '09), May 2009.
[9] W. Yu, S. Chellappan, X. Wang, and D. Xuan, "Peer-to-Peer System-Based Active Worm Attacks: Modeling, Analysis and Defense," Computer Comm.,vol. 31, no. 17, pp. 4005-4017, Nov. 2008.
[10] A. Ganesh, L. Massoulie, and D. Towsley, "The Effect of Network Topology on the Spread of Epidemics," Proc. IEEE INFOCOM, 2005.
[11] Y. Wang, D. Chakrabarti, C. Wang, and C. Faloutsos, "Epidemic Spreading in Real Networks: An Eigenvalue Viewpoint," Proc. IEEE Int'l Symp. Reliable Distributed Systems (SRDS), 2003.
[12] M. Newman, S. Strogatz, and D. Watts, "Random Graphs with Arbitrary Degree Distribution and Their Applications," Physical Rev. E, vol. 64, no. 2, pp. 026118(1-17), July 2001.
[13] R. Pastor-Satorras and A. Vespignani, "Epidemic Dynamics in finite size Scale-Free Networks," Physical Rev. E, vol. 65, no. 3, p. 035108(1-4), Mar. 2002.
[14] O. Diekmann and J. Heesterbeek, "Mathematical Epidemiology of Infectious Diseases: Model Building, Analysis and Interpretation".Wiley,1999.