The International Journal Of Engineering And Science (Ijes)
||Volume||2 ||Issue|| 2 ||Pages|| 214-219 ||2013||
Issn: 2319 – 1813 Isbn: 2319 – 1805

Anticipation Measures For Protecting P2P Networks From Malware Spread

[1,]Tella Ramesh,[2,]T.Sunitha
[1,]*M.Tech (CS) Student,Dept Of CSE,QISCET*
[2,]*Assoc Professor,Dept Of CSE, QISCET*

----------Abstract----------
Malware is the software developed wit-h malicious intentions. The detection of such malware can be done by writing program which can understand the dynamics of malware. Towards end this paper presents an analytical model which can effectively characterize the true nature of malware and how it spreads in peer-to-peer 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.

Keywords *- Malware, peer-to-peer networks, compartmental model.*

Date Of Submission: 19, February, 2013 Date Of Publication: 28, February 2013

I. Introduction

The use of peer-to-peer (P2P) networks as a vehicle to spread malware offers some important advantages over worms that spread by scanning for vulnerable hosts. This is primarily due to the methodology employed by the peers to search for content. For instance, in decentralized P2P architectures such as Gnutella [1] where search is done by flooding the network, a peer forwards the query to it's immediate neighbors and the process is repeated until a specified threshold time-to-live, TTL, is reached. Here TTL is the threshold representing the number of overlay links that a search query travels. A relevant example here is the Mandragore worm [2], that affected Gnutella users. The design of the search technique has the following implications: first, the worms can spread much faster, since they do not have to probe for susceptible hosts and second, the rate of failed connections is less. Thus, rapid proliferation of malware can pose a serious security threat to the functioning of P2P networks. Understanding the factors affecting the malware spread can help facilitate network designs that are resilient to attacks, ensuring protection of the networking infrastructure. This paper addresses this issue and develops an analytic framework for modeling the spread of malware in P2P networks while accounting for the architectural, topological, and user related factors. We also model the impact of malware control strategies like node quarantine.

II. Relationship To Prior Work

Though the initial thrust in P2P research was measurement oriented, subsequent works, [3], [4], [5], have proposed analytical models for the temporal evolution of information in the network. The focus of these works is on transfer of regular files and they do not apply to malware that spread actively.In addition, they are specialized to BitTorrent like networks and cannot be extended for P2P networks such as Gnutella. The issue of worms in peer-to-peer networks is addressed in [6], [7] using a simulation study of P2P worms and possible mitigation mechanisms. Epidemiological models to study malware spread in P2P networks are presented in [8], [9]. These studies assume that a vulnerable peer can be infected by any of the infected peers in the network. This assumption is invalid since the candidates for infecting a peer are limited to those within TTL hops away from it and not the entire network. Another important omission is the incorporation of user behavior. Typically, users in a P2P network alternate between two states: the on state, where they are connected to other peers and partake in network activities and the off state wherein they are disconnected from the network. Peers going offline result in fewer candidates for infection thereby lowering the intensity of malware spread. An empirical model for malware spreading in BitTorrent is developed in [10] while models for the number of infected nodes by

dynamic hit list-based malware in BitTorrent networks is presented in [11], [12]. However, these models ignore node dynamics such as online-offline transitions and are applicable only to BitTorrent networks.In [13], [14], the authors use hypercubes as the graph model for P2P networks and derive a limiting condition on the spectral radius of the adjacency graph, for a virus/worm to be prevalent in the network.The models do not account for the fact that once a peer is infected, any susceptible peer within a TTL hop radius becomes a likely candidate for a virus attack. In the current work, we formulate a comprehensive model for malware spread in Gnutella type P2P networks that addresses the above shortcomings. We develop the model in two stages: first, we quantify the average number of peers within TTL hops from any given peer and in the second stage incorporate the neighborhood information into the final model for malware spread.

III. Malware Propagation Model For P2p Networks

This section presents our framework for modeling malware spread in P2P networks. Our model's focus is on the propagation of malware and not regular files.

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).

$$G_0(x) = \sum_{i=0}^{\infty} p_i x^i$$

The distribution of the degree of a vertex that we arrive at by following an edge from a vertex is different from that of an arbitrary vertex in the graph. An edge arrives at a vertex with probability a proportional to the degree of the vertex. Thus, the probability that a randomly chosen edge leads to a vertex with degree i is proportional to i p_i The pmf of the degree of the vertex can then be obtained from the pmf of an arbitrary vertex by normalizing it with $\sum_i ip_i$ and its probability generating function (pgf) is then

$$\frac{\sum_i ip_i x^i}{\sum_i ip_i} = \frac{xG_0^{'}(x)}{G_0^{'}(1)}$$

TABLE 1
Notation and P2P Model Parameters

$\lambda_{on},\lambda_{off}$	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 delete it
$1/\delta$	Average time for which a peer stores a file

3.2 Compartmental Model

We formulate our model as a compartmental model, with the peers divided into compartments, each signifying it's state at a time instant. In addition, to account for power-law topologies, we develop the compartmental model in terms of the node degree [17]. For each possible node degree k, the network is partitioned into four classes.

TABLE 2
classes in the compartmental model

$P_S^{(k)}$	Number of peers wishing to download a file.
$P_E^{(k)}$	Number of peers currently downloading the malware
$P_I^{(k)}$	Number of peers with a copy of the malware.
$P_R^{(k)}$	Number of peers who either have deleted the malware or are no longer interested downloading any file

Each class has two components: one comprising of peers of that class that are currently online, while the second represents the offline peers. For instance, $P_{I_{on}}^{(k)}$ denotes the peers with degree k infected by the malware that are currently online and $P_{I_{off}}^{(k)}$, the offline infected peers.

Our formulation is based on the principle of mass action, where the behavior of each class is approximated by the mean number in the class at any time instant. By employing the mean-field approach, we make the following assumptions about the system:

- The number of members in a compartment is a differentiable function of time. This holds true in the event of large compartment sizes and since P2P networks comprise of tens of thousands of users, assuming this is quite reasonable.
- By abstracting the P2P graph through differential equations, the emphasis is more on the numbers of each class, rather than the particulars of each member of the respective classes.
- The spread of files in the P2P network is deterministic, i.e., the behavior is completely determined by the rules governing the model. In other words, the properties of a class are dictated by the number of members present.
- The size of the network does not vary over the time during which the spread of malware is modeled.
  We first determine the probability that a susceptible peer with degree k is infected when it tries to download an arbitrary file.
  The dynamics of the spread of malware in peers with degree k can then be represented in terms of the constituent classes by the following deterministic system of equations:

$$\frac{dP_{S_{on}}^{(k)}}{dt} = -\lambda P_{S_{on}}^{(k)}(1-(1-p_{\inf})^{z_{av}^{(k)}}) + r_1 P_{E_{on}}^{(k)}) + r_2 P_{R_{on}}^{(k)} - \lambda_{off} P_{S_{on}}^{(k)} + \lambda_{on} P_{off}^{(k)} \quad (1)$$

$$\frac{dP_{E_{on}}^{(k)}}{dt} = \lambda P_{S_{on}}^{(k)}(1-(1-p_{\inf})^{z_{av}^{(k)}}) - r_1 P_{E_{on}}^{(k)}$$

$$-\mu P_{E_{on}}^{(k)} - \lambda_{off} P_{E_{on}}^{(k)} + \lambda_{on} P_{E_{on}}^{(k)} \quad (2)$$

$$\frac{dP_{I_{on}}^{(k)}}{dt} = \mu P_{E_{on}}^{(k)} - \delta P_{I_{on}}^{(k)} - \lambda_{off} P_{I_{on}}^{(k)} + \lambda_{on} P_{I_{off}}^{(k)} \quad (3)$$

$$\frac{dP_{R_{on}}^{(k)}}{dt} = \delta P_{I_{on}}^{(k)} - r_2 P_{R_{on}}^{(k)} - \lambda_{off} P_{R_{on}}^{(k)} + \lambda_{on} P_{R_{off}}^{(k)} \quad (4)$$

$$\frac{dP_{S_{off}}^{(k)}}{dt} = \lambda_{off} P_{S_{on}}^{(k)} - \lambda_{on} P_{S_{off}}^{(k)} \quad (5)$$

$$\frac{dP_{E_{off}}^{(k)}}{dt} = \lambda_{off} P_{E_{on}}^{(k)} - \lambda_{on} P_{E_{off}}^{(k)} \quad (6)$$

$$\frac{dP_{I_{off}}^{(k)}}{dt} = \lambda_{off} P_{I_{on}}^{(k)} - \lambda_{on} P_{I_{off}}^{(k)} \quad (7)$$

$$\frac{dP_{R_{off}}^{(k)}}{dt} = \lambda_{off} P_{R_{on}}^{(k)} - \lambda_{on} P_{R_{off}}^{(k)} \quad (8)$$

IV. Model Analysis

In this section, we analyze the model presented in the previous section and obtain the expressions governing the global stability of the malware free equilibrium (MFE).

4.1 Malware Free Equilibrium

We now proceed with the derivation of the basic reproduction number, R0, a metric that governs the global stability of the MFE. Here, R0 quantifies the number of vulnerable peers whose security is compromised by an infected host during it's lifetime. It is an established result in epidemiology that $R_0 < 1$ ensures that the epidemic dies out fast and does not attain an endemic state [15]. Stability information of the MFE is important since this guarantees that the system continues to be malware free even if newly infected peers are introduced.
We follow the methodology presented in [16], [17], where "next generation matrices" have been proposed to derive the basic reproduction number. In this method, the flow of peers between the states are written in the form of two vectors F and V. The ith element of F is the rate of appearance of new infections in compartment i and the ith element of V is defined as $V_i = V_i^- - V_i^+$, where V_i^+ is the rate of transfer of peers into compartment i by all other means and V_i^- is the rate of transfer of peers out of compartment i. These vectors are then differentiated with respect to the state variables, evaluated at the malware free equilibrium, and only the part corresponding to the infected classes are then kept to form the matrices F and V, i.e.,

$$F = \left[\frac{\partial F_i}{\partial x_j}(x_0)\right], V = \left[\frac{\partial V_i}{\partial x_j}(x_0)\right], 1 \le i, j \le m \qquad (9)$$

Where F_i and V_i are the ith entries of F and V.

4.2 Quarantine

As a form of damage control, the intensity of malware spread can be limited by quarantining infected nodes. This section quantifies the impact of the quarantine rate on the basic reproduction ratio R_0. Quarantine is introduced in the system as follows: we assume that an infected node is taken off the network with probability η. We also assume that this operation does not result in the P2P network being split into disconnected components. The quarantined peers comprise a new compartment $P_Q^{(k)}$ and when rid of malware, enter the recovered state at rate . This introduces the following changes to the system of (1-8):

- Additional terms to the classes $P_{I_{on}}^{(k)}$ and $P_{R_{on}}^{(k)}$ reflecting the departure of quarantined peers and addition of recovered peers, respectively.
- An additional equation describing the evolution of $P_Q^{(k)}$

The following equations represent additional terms.

$$\frac{dP_{I_{on}}^{(k)}}{dt} = \mu P_{E_{on}}^{(k)} - \delta P_{I_{on}}^{(k)} - \lambda_{off} P_{I_{on}}^{(k)} + \lambda_{on} P_{I_{off}}^{(k)} - \eta P_{I_{on}}^{(k)} \quad \frac{dP_{R_{on}}^{(k)}}{dt} = \delta P_{I_{on}}^{(k)} - r_2 P_{R_{on}}^{(k)} - \lambda_{off} P_{R_{on}}^{(k)} + \lambda_{on} P_{R_{off}}^{(k)} + \vartheta P_Q^{(k)}$$

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

$$\frac{dP_Q^{(k)}}{dt} = \eta P_{I_{on}}^{(k)} - \vartheta P_Q^{(k)}$$

V. 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.

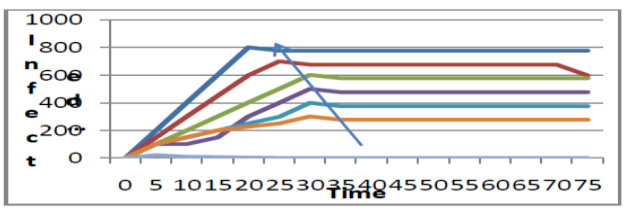

Fig.1: Effect of λ on malware intensity

Fig.1 shows the results which visualize the time and infected peers. When time grows, the benefits of offline users also more.

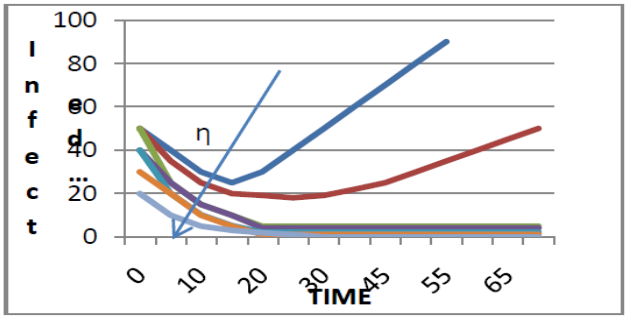

Fig.2: Effect of quarntine on malware intensity

As can be seen in fig. 2, 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.

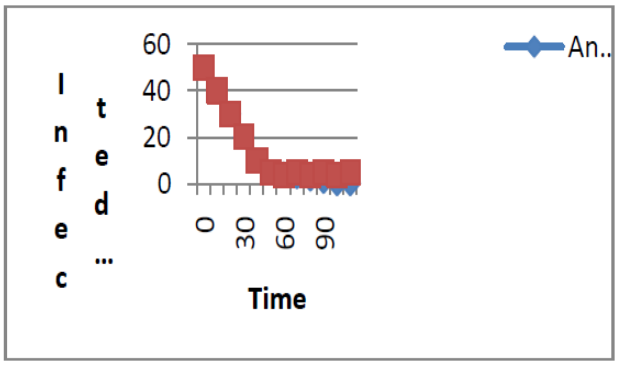

Fig. 3: Impact on malware intensity(λ =0.005)

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 $\Re_0$ is less than 1, the

Number of infected peers is dropping down to zero.

VI. CONCLUSION

In this paper, we developed an analytic model to understand the dynamics of malware spread in P2P networks. The need for an analytic framework incorporating user characteristics (e.g., offline to online transitional behavior) and communication patterns (e.g., the average neighbourhood size) was put forth by quantifying their influence on the basic reproduction ratio. It was shown that models that do not incorporate the above features run the risk of grossly overestimating $\Re_0$ and thus falsely report the presence of an epidemic.

References

[1] Clip2, "The Gnutella Protocol Specification v0.4," http://www.clip2.com/ GnutellaProtocol04.pdf, Mar. 2001.

[2] E. Damiani, D. di Vimercati, S. Paraboschi, P. Samarati, and F. Violante, "A Reputation-Based Approach for Choosing Reliable Resources in Peer-to-Peer Networks," Proc. ACM Conf. Computer and Comm. Security (CCS), pp. 207-216, Nov. 2002.

[3] X. Yang and G. de Veciana, "Service Capacity in Peer-to-Peer Networks," Proc. IEEE INFOCOM '04, pp. 1-11, Mar. 2004.

[4] D. Qiu and R. Srikant, "Modeling and Performance Analysis of BitTorrent-Like Peer-to-Peer Networks," Proc. ACM SIGCOMM, Aug. 2004.

[5] J. Mundinger, R. Weber, and G.Weiss, "Optimal Scheduling of Peer-to-Peer File Dissemination," J. Scheduling, vol. 11, pp. 105-120, 2007.

[6] 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.

[7] 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.

[8] F. Wang, Y. Dong, J. Song, and J. Gu, "On the Performance of Passive Worms over Unstructured P2P Networks," Proc. Int'l Conf. Intelligent Networks and Intelligent Systems (ICINIS), pp. 164-167, Nov. 2009.

[9] R. Thommes and M. Coates, "Epidemiological Models of Peer-to-Peer Viruses and Pollution," Proc. IEEE INFOCOM '06, Apr. 2006.
[10] 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.
[11] 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.
[12] 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.
[13] A. Ganesh, L. Massoulie, and D. Towsley, "The Effect of Network Topology on the Spread of Epidemics," Proc. IEEE INFOCOM, 2005.
[14] 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.
[15] O. Diekmann and J. Heesterbeek, Mathematical Epidemiology of Infectious Diseases: Model Building, Analysis and Interpretation. Wiley, 1999.
[16] P. van den Driessche and J. Watmough, "Reproduction Numbers and Sub-Threshold Endemic Equilibria for Compartmental Models of Disease Transmission," Math. Biosciences, vol. 180, pp. 29-48, 2002.
[17] J. Arnio, J. Davis, D. Hartley, R. Jordan, J. Miller, and P. van den Driessche, "A Multi-Species Epidemic Model with Spatial Dynamics," Math. Medicine and Biology, vol. 22, pp. 129-142, Mar. 2005.

AUTHORS PROFILE

TELLA RAMESH is a young auther and student of M. Tech in the stream of Computer Science at QIS College of Engineering and Technology which is affiliated under the university of JNTU, Kakinada.

T. SUNITHA Associate Professor, Department of Computer science engineering, QIS College of Engineering and Technology which is affiliated under JNTU, Kakinada which is permanent NBA accredited Institute. She studied Btech in S.S.N enggg college, Ongole, Mtech in Andhra University. She has 7 years of teaching experience.