ISSN:0975-9646

Spreading of Malware Dynamically in Peer To Peer Network

J. Sushma

Department of computer science and engineering,
Lords Institute Of Engineering & Technology, JNTUH ,
Hyderabad , AP,INDIA

Abstract— **In this paper, we formulate an analytical model to characterize the spread of malware in decentralized, Gnutella type peer-to-peer (P2P) networks and study the dynamics associated with the spread of malware. Using a compartmental model, we derive the system parameters or network conditions under which the P2P network may reach a malware free equilibrium. The model also evaluates the effect of control strategies like node quarantine on stifling the spread of malware. The model is then extended to consider the impact of P2P networks on the malware spread in networks of smart cell phones.**

Keywords—**Malware propagation, peer-to-peer networks, Internet worms and viruses.**

I. Introduction

Peer to peer network (p2p) was used as a vehicle for spreading up of Malware which offers few important advantages over worms spread by scanning for vulnerable host was due to the primary method utilized by the peers to search the content. Such as for Gautella decentralized P2P architecture search is done by flooding of network.[1]

A peer forwards the query to its next immediate neighbor and the process is continued till the specified threshold TTL (time to live) is reached where TTL represents the number of overlay links of search query travels for an example Mandragore work that affects the Gnuteller user. Here whenever the user searches for a medial file in an infected computer the virus appears as a result for every request and the user thinks that the affected file is the result he searched for. The design implication for the search technique is as follows. First one is worms can spread faster because they don't have problem for susceptible hosts & the second one is the rate of failed connection is less. Thus Malware spreads serious security threat to P2P network.[2]

If the factors affecting the malware spread is understood we can modify the network design in order to project the networking infrastructure. The paper consists of issues and developments of an analytic frame work for modeling the spread of malware in peer to peer network and also the impact of control strategies like node quarantine.

II. Propagation Model Of Malware For Peer To Peer Network.

Models are focused on propagation of malware but not on regular files.

A. Search Mechanism:

Initiation for transfer of information in peer to peer network is done by search request. Search mechanism is employed by flooding in Gnutella networks. In order to search for a failed a query is forwarded to its neighbors. After receiving the query it takes the responds and checks the TTL query. If the value is greater than zero it forwards the query else its discarded. The model malware spread is imperative to determine the average rate where the query reach a mode which depends on search neighbor.

Generating function to quantify the search neighborhood is defined by the probability function of vertex degree:

$$G0(x)=\sum_{i=0}^{\infty} P_i x^i$$

Where P_i is probability of randomly chosen vertex having degree i.

Gnutella network has a power law distribution Therefore $P_i=C_i^{-\tau}$ where C and τ are constant[3]

The heterogeneity of the connectivity distribution affects the search nodes with different degrees. Therefore we evaluate the neighborhood size of vertex as a function of its degree k.

The distribution of a degree of the vertex we arrive from an edge is different from arbitrary vertex in the graph. The probability that a randomly choose vertex with degree i is proportional to ip_i.

The probability mass function of the degree of the vertex can be obtained from the probability mass function of an arbitrary vertex by normalizing it with $\sum_i ip_i$ and its probability generating function.

$$\frac{\sum_i ip_i x^i}{\sum_i ip_i} = \frac{xG'_0(x)}{G'_0(1)}$$

Randomly chosen edge to reach vertex is continued till we reach all m hop neighbors. The numbers of vertices arrived has degree distribution above, less one power of x to compensate to edge we arrive.

The probability generating function for outgoing edges is given by :

$$G_1(x) = \frac{G_0'(x)}{G_0'(1)}$$

With N nodes in the network the probability of the outgoing edges connecting to the original vertex any of its immediate neighbors falls to N^{-1} and can be neglected as $N \rightarrow \infty$

For a 2-hop neighbors the probability generating function is given by

$\sum_k p_k|G_1(x)|^k = G_0(G_1(x))$

Similarly for m-hop neighbor its given by $G_0(G_1(G_1(.....G_1(x))))$with m-1 iterations

If the node has a degree k the probability of generating function of its degree is given by

$G_0^{(k)}(x) = x^k$

In terms of recursive convolution the probability generating function for m-hop neighbors of a node with k degree is defined by

$G_m^{(k)}(x) = x^k$ for m=1

$G_m^{(k)}(x) = (G_1(G_1(....(G_1(x))))$ for $m \geq 2$

differentiating and substituting x=1 the average number of m hop neighbors is given by

$$z_m^{(k)} = \frac{dG(m)}{dx}\Big|_{x=1} = G_0^{(k)\prime}(1)[G_1'(1)]^{m-1}$$

$$= K\left[\frac{z_2}{z_1}\right]^{m-1}$$

Where $Z_2 = G_0''(1)$ and $Z_1 = G_0'(1)$ since the search of the neighborhood of peer extends up to TTL hops & the average neighborhood size is given by

$$z_{av}^{(k)} = \sum_{i=1}^{TTL} z_i^{(k)} = K\frac{z_1}{z_2 - z_1}\left[\left(\frac{z_2}{z_1}\right)^{TTL} - 1\right]$$

B. *Compartmental Model:*

The peers are divided into compartments each signifying its state at an instant time. In addition of power law topologies , we develop a model in terms of mode degree. For a node degree k the network is partitioned into four classes as below:

(i) PS(K): Number of peers wishing to download a file.

(ii) PE(K): Number of peers currently downloading the malware

(iii) PL(k): Number of peers with copy of the malware.

(iv)PR(K): Number of peers either deleted the malware or are no longer interested downloading any file.

Each class has two components:

1) Consists of peers that are online.

2) Consists of peers that are offline.

P1(K) denotes peer with degree k infected by online malware

P1 off(K): denotes peer with degree is infected by off line malware

Since network consists of finite nodes it consists of finite classes. Np(K) is the total number of nodes with degree k for both one and off line.

table -I: notation and peer to peer 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 deleted it
$1/\delta$	average time for which a peer stores a file

The above table parameters are used in our model .

The following assumptions are employed by mean field approach.

- The number of members in a compartment is a differentiable function of time. This holds true in the event of large compartment sizes & 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 rule governing the model.
- The size of network does not vary over the time during which the spread of malware is modeled.

The probability of peer is infected when tried to download arbitrary file. The probability of a neighbor of an arbitrary mode with degree j is given by $\frac{jp_i}{\bar{Z}}$,where

$$\bar{z} = \sum_i ip_i$$

When query reach the node with degree j, its infected respond to the query is given by the probability $p_{1_{on}}^{(i)} / N_p^{(j)}$.

Therefore, the probability of infected arbitrary neighbor is given by:

$$p_{\inf} = \sum_j \frac{jp_j}{\bar{z}} \frac{p_{1_{on}}^{(j)}}{N_p^{(j)}}$$

Search initiated by a node average reaches Z peer. Then the probability of at least one of the peer responds to the query & gets infected is

$$\left(1-(1-p_{\inf})^{z_{av}^{(k)}}\right)$$

Dynamic spread of malware in peer is represented in terms of constituent classes by below equations.

$$\frac{dp_{s_{on}}^{(k)}}{dt} = -\lambda p_{s_{on}}^{(k)}\left(1-(1-p_{\inf})^{z_{av}^{(k)}}\right) + r_1 p_{E_{on}}^{(k)} + r_2 p_{R_{on}}^{(k)}$$

$$-\lambda_{off}\, p_{s_{on}}^{(k)} + \lambda_{on} p_{s_{off}}^{(k)}$$

$$\frac{dp_{I_{on}}^{(k)}}{dt} = -\mu p_{E_{on}}^{(k)} - \delta p_{I_{on}}^{(k)} - \lambda_{off}\, p_{I_{on}}^{(k)} + \lambda_{on} p_{I_{off}}^{(k)}$$

$$\frac{dp_{s_{off}}^{(k)}}{dt} = \lambda_{off}\, p_{s_{on}}^{(k)} - \lambda_{on} p_{s_{off}}^{(k)}$$

$$\frac{dp_{I_{off}}^{(k)}}{dt} = \lambda_{off}\, p_{I_{on}}^{(k)} - \lambda_{on} p_{I_{off}}^{(k)}$$

(k) occurs if peer goes offline or initiates search query that is successful. The former occurs at rate off while the latter is contingent on the rate at the request for file download are generated, multiplied by the probability the query reaches at least one infected node in online state. The rate at which transitions from $p_{s_{on}}^{(k)}$ into $p_{E_{on}}^{(k)}$ occurs is given by $\lambda p_{s_{on}}^{(k)}\left(1-(1-p_{\inf})^{z_{av}^{(t)}}\right)$. The membership of class $p_{s_{on}}^{(k)}$ increases if the reasons are as follows.

- An offline peer of class $p_s^{(k)}$ comes online.
- A peer currently downloading terminate the process, say due to unsatisfactory download speeds.
- A peer that previously had the file, either accidentally or intentionally deletes the file& wishes to download it again.

The peers per unit time exist class $p_{s_{on}}^{(k)}$ total is: $\left(\lambda_{off} + \lambda\left(1-(1-p_{\inf})^{z_{av}^{(k)}}\right)\right)p_{s_{on}}^{(k)}$

The peer per unit time entering the class of number $r_1 p_{E_{on}}^{(k)} + r_2 p_{R_{on}}^{(k)} + \lambda_{on} p_{S_{off}}^{(k)}$

When both entering & existing is combined it gives the rate of change of membership of class $p_{S_{on}}^{(k)}$. The model presented above represents an upper bound for the infected nodes because correlation in the neighborhood node is within TTL hops are neglected.

Since malware size ranges to few kilo bites the download time is expected to be small than on off transmission times. Thus mean field approximation are accepted.

III. Model Analysis

A. *Malware Of Equilibrium*

Basic reproduction number Ro quantifies the number of vulnerable peers whose security is compromised by an infected in its lifetime. If R_0<1. epidemic dies fast and does not attain an endemic state. Stability of Malware free Equilibrium is important because it resembles that the system is malware free even if newly infected peers are introduced.

In the next generation matrices methodology the flow of peers between the state are written in the form of vector F and V. The ith element of F is the rate of appearance of new infection in the ith compartment and the ith element of V is defined as $V_i = V_i^- - V_i^+$ where Vi is the rate of transfer of peers into ith compartment and Vi - is the rate of transfer of peers out of the ith compartment. They are then differentiated with respect to State variables.

$$F = \left[\frac{\partial F_i}{\partial x_j}(x_0)\right],\ V = \left[\frac{\partial V_i}{\partial x_j}(x_0)\right]$$

1≤i , j≤m, $x_i = F_i(x) - V_i(x)$,(x_0) is malware free equilibirum and m is the number of infectious state.

The basic reproduction number, Ro, is then the largest absolute eigen value (special radius)p(), of the matrix FV^{-1}, i.e. $R_0 = \rho\left(FV^{-1}\right)$. Using elementary matrix algebra and rearranging the terms, it can be easily verified that the a product FV^{-1} can be broken down into GB^{-1}CA^{-1}, with the constituent matrices as enumerated above. Thus,

$$R_0 = p\left(\mathrm{GB}^{-1}\mathrm{CA}^{-1}\right)$$

B. *Quarantine:*

For a 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 Ro. Quarantine is introduced in the system as follows: we assume that an infected node is taken off the network with probability n. We also assume that tahis 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 ϑ This introduces the following changes to the sytem.

Additional terms to the classes $p_{I_{on}}^{(k)}$ and $p_{R_{on}}^{(k)}$ reflecting the departaurae of quarantied peers and addition of recovered peers, respectively.

An additional equation describing the evolution of $p_Q^{(k)}$. Thus they are, rspectively, modified to.

$$\frac{dp_{I_{on}}^{(k)}}{dt} = \mu P_{E_{on}}^{(k)} - \delta p_{I_{on}}^{(k)} - \lambda_{off}\, p_{I_{on}}^{(k)} + \lambda_{on}\, p_{I_{off}}^{(k)} - \lambda p_{I_{on}}^{(k)}$$

$$\frac{dp_{R_{on}}^{(k)}}{dt} = \delta p_{I_{on}}^{(k)} - r_2\, p_{R_{on}}^{(k)} - \lambda_{off}\, p_{R_{on}}^{(k)} + \lambda_{on}\, p_{R_{off}}^{(k)} + \vartheta p_Q^{(k)}$$

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

$$\frac{dp_Q^{(k)}}{dt} = \eta p_{I_{on}}^{(k)} - \vartheta p_Q^{(k)}$$

the relevant matrices for computing R$_0$ are modified as

$$F = \begin{bmatrix} 0 & G & \tilde{0} \\ 0 & 0 & \tilde{0} \\ \tilde{0}^{\tau} & \tilde{0}^{\tau} & \hat{0} \end{bmatrix}, \quad v = \begin{bmatrix} A & 0 & \tilde{0} \\ -c & \tilde{B} & \tilde{0} \\ \tilde{0} & D & E \end{bmatrix}$$

$$\tilde{B} = \begin{bmatrix} \lambda + \delta + \lambda_{off} & . \; . & 0 & 0 & . \; . & 0 \\ . & . \; . & . & . & . \; . & . \\ 0 & . \; . & .\lambda + \delta + \lambda_{off} & 0 & . \; . & 0 \\ 0 & . \; . & 0 & \lambda_{on} & . \; . & 0 \\ 0 & . \; . & 0 & 0 & . \; . & \lambda_{on} \end{bmatrix} - \tilde{M}$$

$$D = \begin{bmatrix} -\eta & . & . & 0 & 0 & . & . & 0 \\ . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & . & . \\ 0 & . & . & -\eta & 0 & . & . & 0 \end{bmatrix} E = \begin{bmatrix} \vartheta & . & . & 0 \\ . & . & . & . \\ . & . & . & . \\ 0 & . & . & \vartheta \end{bmatrix}$$

$$R_V = \rho\left(FV^{-1}\right)$$

$$\frac{dp_{s_{on}}}{dt} = r_2 P_R - \lambda P_{S_{on}} \left[1 - \left[1 - \frac{p_1}{N_p} \right]^{Z^{av}} \right] - \lambda_{offP_{s_{on}}} + \lambda_{on} P_{s_{off}}$$

$$\frac{ds_{off}}{dt} = \lambda_{off}\, p_{s_{on}} - \lambda_{on}\, p_{s_{off}}$$

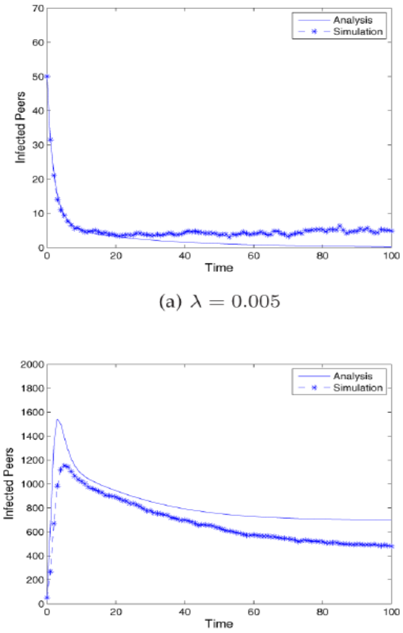
$$FV^{-1} = \frac{1}{\vartheta(\delta + \eta)} \begin{bmatrix} \vartheta \lambda z_{av}\, p_{on} & 0 \\ 0 & 0 \end{bmatrix}$$

$$R_0 = \frac{\lambda z_{av}\, p_{on}}{(\delta + \eta)}$$

The malware spread does not reach epidemic proportions provided R$_0$<1 and hence the required rate for quarantining infected peers need is $\eta > \lambda z_{av} p_{on} - \partial$. such a measure takes the node of the peer to peer network and thus it would not be able to participate in any future file transfers until the malware has been completely removed. this is indeed necessary since an infected peer always responds positively to any query with the malware cloaked as the file being searched for. Thus, the only way to prevent the node from infecting other is to take it off the network.

IV. Conclusion:

The simulations were conducted using a custom bit simulator. Results are reported for a 10000 node network with a power law graph topology with τ =3.4

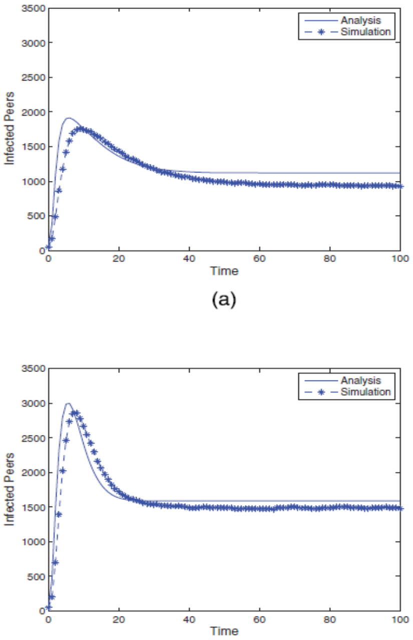


(a) $\lambda = 0.005$

(b) $\lambda = 2.0$

Figure 1: impact of λ on malware intensity for the system.

The above figure sustain our analytical results that requires basic reproduction. The basic figure 2.a and 2.b are observed for off line duration on malware intensity.



(a)

(b)

Figure 2: offline duration of malware intensity

REFERENCE:

[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] 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.