A Reputation Based Approach For Choosing Reliable Resources In Peer To Peer Networks

E. Ravichandra Reddy, V. Madhavi

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. The need for an analytic framework incorporating user characteristics (e.g., offline to online transitional behavior) and communication patterns (e.g., the average neighborhood size) was put forth by quantifying their influence on the basic reproduction ratio. 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.

Index Terms—malware Propagation, modeling, Peer-to-Peer networks.

1. 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 its 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. 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.

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.

2. RELATED WORK

2.1. Epidemic Spreading in Real Networks: An Eigenvalue Viewpoint

Computer viruses remain a significant threat to today's networks and systems. Existing defense mechanisms typically focus on local scanning of virus signatures. While these mechanisms can detect and prevent the spreading of known viruses, they do little for globally optimal defenses. The recent proliferation of malicious code that spreads with virus code exacerbates the problem. From a network dependability stand point, the propagation of malicious code represents a particular from of fault propagation, which may lead to the ultimate demise of the network (considering distributed denial-of- service attacks). With the exception of a few specialized modeling studies, much still remains unknown about the propagation characteristics of computer viruses and the factors that influence them.

2.2. The impact of human mobility on spatial disease dynamics

Understanding human mobility is crucial for modeling the spatial spread of human infectious diseases. The quantitative description of spatial epidemics is based on two prominent theoretical approaches, diffusive dispersal and direct coupling or effective force of infection. The first ansatz assumes random walk movement of the host between different locations whereas the second employs an effective force of infection between distinct populations. Both models are inconsistent with important aspects of human mobility, most importantly the bidirectional movements between individuals' homes and distant location. We investigate a novel epidemiological model that explicitly takes into account this bidirectional nature of human

movements. In various topologies (networks and lattices) we find significant differences as well as similarities among all three models, depending on the parameters. On a lattice we obtain an analytical expression for the velocity of the propagating epidemic front. In contrast to the diffusion approach, our model predicts a saturation of the velocity with increasing traveling rate. Our analysis is supported by numerical simulations on both lattices and networks and provides a framework for incorporating the abundance of pervasive data on individual human mobility into disease dynamics modeling.

Infectious diseases remain a pressing challenge for mankind [1]. Investigations on epidemics in human and animal populations require accurate assessments of their spatiotemporal dynamics [2] as infectious diseases spread among different locations due to movements of their host. It is frequently assumed that hosts move chaotically or perform random walks in space and therefore host mobility has been described within reaction-diffusion dynamics.

Here we propose a method for incorporating bidirectional mobility as random movements on overlapping individual topologies. In particular we consider star-shaped network topologies corresponding to commuting movements of individuals between their home location (center node) and accessible destinations (distant nodes) and investigate properties of the resulting epidemics on regular lattices and random networks.

Individuals moving on these star-shaped networks are required to return to their home location from a visit to a distant location before they visit another distant location.

2.3. Propagation Modeling of Passive Worms in P2P Networks

Studying worm propagation using the aggregated properties of P2P networks typically assumes a static topology, in which a node stores the addresses of all neighbors with which it had communicated. The lack of detailed peer interactions makes topology-driven models unsuitable to simulate worm propagation, for instance, if a node can only cache the last $n\psi$ communicating peers. It is also difficult, if not impossible, to use these models to study passive P2P worms. P2P networks are complex systems and it may not be feasible to use an analytical approach to model worm propagations without making overly simplified assumptions. Instead we present a unified simulation framework, driven by a P2P file-sharing workload model, to study the passive P2P worms. Unlike previous work, our approach models detailed peer-to-peer file-sharing interactions. Our model captures file requests and downloads, which lead to network activities and topologies; thus it can be used to study the propagations of passive worms.

Considering the patterns of worm propagation are different at different stages of worm propagation, we model P2P passive worm propagation at different stages separately. To model in the mean-filed method [4], it is necessary to explain these parameters and assumptions employed in the following models.

A. Model Parameters and Assumptions

The intent of our model is to predict the expected behavior of a worm which spreads through a P2P network in the form of malicious code embedded in executable files shared by peers. We make the simplifying assumption as follows.

1) Each user put all files, which can be downloaded by others, to his/her shared folder. And all users download files to their shared folder. Peers online refer to those P2P clients which are running.

2) The number of peers online is invariable. In this situation, no peers added or exited, and no new files are added.

3) After downloaded, a file is executed at once.

4) Time spent on searching, connecting, downloading and executing a file, is invariable, which is call as a time unit. It takes a time unit that an infected peer returns to the susceptible state or is immunized.

5) When a peer is infected, c infected files reside the peer's shared folder and have c different names. All infected peers share the same c infected files.

We are not concerned with the transfer of media files which cannot contain malicious code, and do not model them. Note that we use the term user in this paper to refer to a person using a P2P client program. The term peer is used to collectively refer to a P2P client and the user directing its behavior. In order to formally analyze attack strategies and epidemiological modeling of P2P worms, we list

the most parameters in table 1, which will have an impact on worm attack effects.

TABLE I. NOTATIONS IN MODELS

$N(t)$	Number of all hosts on the P2P network at time t, here it is a constant.
$S(t)$	Number of susceptible hosts at time unit t.
$I(t)$	Number of infected hosts at time unit t.
$R(t)$	Number of recovered host at time unit t.
$K(t)$	Number of infected files at time unit t.
$M(t)$	Number of uninfected files at time unit t.
$h(t)$	Possibility of downloading an infected file at time unit t, $h(t) = \alpha \frac{K(t)}{M(t)+K(t)}$,
λ_d	Average rate, in files per time unit, at which each peer downloads new files (this includes time spent searching, setting up the connection to another peer and executing download files.
λ_{is}	Average rate, in hosts per time unit, at which infected hosts return to susceptible hosts.
λ_{sr}	Average rate, in removes per time unit, at which susceptible hosts are immunized.
λ_{ir}	Average rate, in removes per time unit, at which infected hosts are immunized.

3. MODEL FOR P2PNETWORKS

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

The transfer of information in a P2P network is initiated with a search request for it. This paper assumes that the search mechanism employed is flooding, as in Gnutella networks. In this scenario, a peer searching for a file forwards a query to all its neighbors. A peer receiving the query first responds affirmatively if in possession of the file and then checks the TTL of the query. If this value is greater than zero, it forwards the query outwards to its neighbors, else, the query is discarded. In our scenario, it suffices to distinguish any file in the network as being either malware or otherwise. This is because, as noted earlier, an infected peer replies affirmatively to all the queries that it receives with the malware being substituted for the file being searched for. Thus to model malware spread, it is imperative to determine the average rate at which queries reach a node, which in turn depends on the search neighborhood.

3.2 Analytical Model

We formulate our model as a compartmental model, with the peers divided into compartments, each signifying its state at a time instant. In addition, to account for power-law topologies, we develop the compartmental model in terms of the node degree [7]. For each possible node degree k, the network is partitioned into four classes:

- Ps(k): Number of peers wishing to download a file.
- PE(k): Number of peers currently downloading the malware.
- PI(k): Number of peers with a copy of the malware.
- PR(k): Number of peers who either have deleted the malware or are no longer interested downloading any file.

Further, 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, PIon (k) denotes the peers
With degree k infected by the malware that are currently online and PIoff (k), the offline infected peers. Note that since we consider networks with a finite number of nodes, the number of classes is finite, even with power-law topologies. We denote by Np the total numbers of peers in the network and by Np (k) the total number of nodes with degree k, both online and offline. Table 1 defines the parameters used in our model.

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.

The model presented above represents an upper bound on the number of infected nodes. This is because the model neglects the correlations in the neighborhoods of nodes that are within TTL hops of each other. Also, since malware sizes are typically small (less than a few kilobytes), the download times are expected to be smaller than the on-off transition times of peers which are of the order of hours. Thus, the mean-field approximations used in our analysis are acceptable.

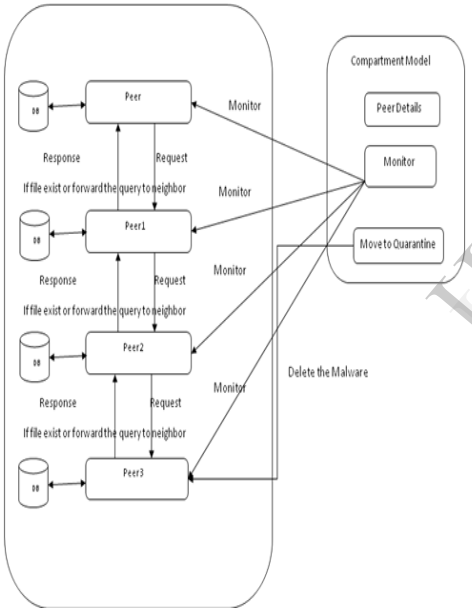

Fig 1. Malware attack process

4 MODEL ANALYSES

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 follow the methodology presented in [7], [10], where "next generation matrices" have been proposed to derive the basic reproduction number. In this method, the flow of peers between the states is written in the form of two vectors F and V. The *i*th element of F is the rate of appearance of new infections in compartment i and the *i*th 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.,

$$F = \left[\frac{\partial \mathcal{F}_i}{\partial x_j}(x_0)\right],\ V = \left[\frac{\partial \mathcal{V}_i}{\partial x_j}(x_0)\right],\ 1 \le i, j \le m,$$

Where F*i* and V*i* are the *i*th entries of F and V, xi is the *i*th system state variable with xi = F*i*(x)—V*i*(x), (xo) is the malware free equilibrium and m is the number of infectious states. For calculating F and V, the column vectors F and V may be considered to consist of m rows, each corresponding to an infectious state

4.3. 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 R0. Quarantine is introduced in the system as follows: we assume that an infected node is taken off the network with probability. We also assume that this operation does not result in the P2P network being split into disconnected components.

5. CONCLUSIONS

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 neighborhood 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 R0 and thus falsely report the presence of an epidemic.

6. ACKNOWLEDGMENTS

We would to thank the anonymous referee for helpful Comments.

7. REFERENCES

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

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

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

[4] 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.

[5] 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.

[6] 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.

[7] R. Thommes and M. Coates, "Epidemiological Models of Peer-to-Peer Viruses and Pollution," Proc. IEEE INFOCOM '06, Apr. 2006.

[8] J. Schafer and K. Malinka, "Security in Peer-to-Peer Networks: Empiric Model of File Diffusion in Bit Torrent," Proc. IEEE Int'l Conf. Internet Monitoring and Protection (ICIMP '09), pp. 39-44, May 2009.

[9] 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.

[10] 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.

[11] A. Ganesh, L. Massoulie, and D. Towsley, "The Effect of Network Topology on the Spread of Epidemics," Proc. IEEE INFOCOM, 2005.

[12] 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.

Author Bibliography

Mr. E. Ravichandra Reddy received his B.Tech in Computer science and Engineering from Vijay Rural Engg. College, JNTU, Hyderabad and Pursuing M.Tech in Computer science Engineering from Aurora's Technological And Research Institute, JNTU, Hyderabad.
Email:-- eravichandrareddy@gmail.com

Mrs. K. Madhavi working as Sr.Assistant Professor in the Department of Computer Science and Engineering, in Aurora's Technological And Research Institute with a teaching experience of 5 years. She has received her Master Degree in Technology in computer science From VNR VJIET JNTU, Hyderabad.
Email:-- madhavi.vagu@gmail.com