Propagation Modeling of Passive Worms in P2P Networks

Chaosheng Feng, Zhiguang Qin
School of Computer Science & Engineering
University of Electronic Science and Technology of China
Chengdu, China

Laurence Cuthbet, Laurissa Tokarchuk
Department of the Electronic Engineering
Queen Mary, University of London
London, UK

Abstract—**Recent years, researchers have recognized the damage resulting from P2P worms, and some works on P2P worms have been done. However, compared with the study of active worm propagation, passive worm propagation has been less highlighted. Passive worms propagate slowly in Internet, but P2P system can be a potential vehicle to fast the propagation of passive worms. In this paper, we address the issue by analyzing the passive worm propagation models in P2P networks and passive worm propagation is modeled in the mean-field method. The fact that the theory values are in consistence with the simulation values shows that these models proposed are valid and can be used to analyze and predict P2P worm propagation patterns.**

Keywords—**passive worms, modeling, propagation, simulation P2P networks**

I. INTRODUCTION

Millions of Internet users are using large-scale peer-to-peer (P2P) networks to share content files today. Listed on the website of Kazaa, a popular P2P software, there have been almost 39 millions downloads in total and more than 0.8 millions downloads in a single week (November 14, 2005) [1]. During the one-month period of November 2001, Staniford et al. observed 9 million distinct remote IP addresses engaged in successful Kazaa connections with hosts in a single university (about 5,800 Kazaa hosts). The eDonkey2000 network alone typically has over 2 million users connected at any given time [2], while the number of users of the BitTorrent[3] , a most popular P2P file-transferring system , is more than 10 millions.

The widely-deployed P2P systems used by end users, however, have strong security implications. First, the users may have downloaded files embedded with malicious code. Second, the P2P client software may contain vulnerabilities that could be exploited by attackers. In particular, P2P systems often have homogeneous client implementation. For example, a recent study found that more than 75% Gnutella clients run the same software[4]. A single implementation weakness of a commonly used P2P client thus results in a large vulnerable population. This situation is attractive for adversaries to exploit the P2P networks using Internet worms, which can automatically propagate through the network using a single vulnerability without human intervention. The compromised P2P nodes may be used to capture end users' sensitive information or be used for further attacks [5].

In this paper, we focus on studying unstructured P2P file-sharing networks such as BitTorrent and eDonkey. Most worms target these networks. Our purpose is to model the epidemic spreading of passive worms in P2P networks. In next section, we will introduce passive worms in detail. This paper contributes as follows.

1) Propose three models of passive worm propagation, which are suitable for different stages of worm propagation, respectively.
2) Use the numerical analysis tool, Matlab Simulink, to work out the theory values of these models.
3) Develop a simulation frame based on the P2P simulation platform Peersim to simulate such realistic P2P networks as BitTorrent and eDonkey.
4) Verify the validity of the three models proposed by comparing theory values with simulation values.

The rest of this paper is organized as follows. We simply introduce the existing studies of worm propagation in Section II. In Section III, we present three models of passive worms and simply address their relation. In Section IV, we evaluate the performance of the three models by comparing theory results with simulation results. Finally we conclude and point out the future work in Section V.

II. RELATED WORK

A. Key Features of P2P Networks and P2P Passive Worms

This section highlights the key features shared by popular P2P Networks, including BitTorrent, eDonkey2000, and Gnutella [6]. Every peer connected to the network has a shared folder containing all the files the user wishes to make publicly available for download by others on the network. When a user wants to download a file, he begins by sending out a search request. Eventually he will receive back a list of files matching the search criteria. The specific manner in which this list is generated varies among the various P2P networks, but in all cases the query response is the result of the examination of the shared folders of a subset of all peers connected to the network. Once the user elects to download one of the files from the list, his client attempts to set up a connection to a

peer sharing the file and begins receiving the file. Depending on the specific network, the client may attempt to simultaneously download different parts of the file from a number of peers in order to expedite the operation. P2P clients typically save new downloaded files in the shared folder – making them immediately available to other users.

A number of passive worms that exploit P2P networks have already surfaced. The majority of these behave in a similar fashion. Specifically, when a user downloads a file containing the worm and executes it, a number of new files containing the worm are created and placed in the client's shared directory. Some types of viruses, including Achar [7] and Gotorm [8], generate a fixed list of filenames when executed. More advanced viruses, such as Bare [9] and Krepper [10], randomly pick the list of filenames from a large pool of candidates.

B. Existing Modeling Work on Worms

Modeling and analysis of the propagation of worms have been studied for several years. Staniford et al. used the classical simple epidemic model to model the spread of Code Red worm [1]. Zou et al. presented two-factor worm model that considered human countermeasures and network congestion effect [11].Chen et al. presented discrete-time version worm model that considered patching and cleaning effect [12]. Staniford et al. presented the "hit-list worm" and "flash worm" [1]. "Routing worm" can greatly reduce worm's scanning space and fast worm propagation [13]. Staniford et al. presented the concept of contagion worm, which is a passive worm [1]. They also concluded that P2P system is well suitable for contagion worm propagation, but they didn't give detailed modeling and analysis. Yu et al. researched active worm propagation on top of P2P systems [14,15,16]. Existing work on P2P worms has focused on proactive worms that propagate using network topologies [12,17], given the empirical evidence that the P2P topologies approximate power-law distributions [18,19].

III. P2P PASSIVE WORM PROPAGATION MODELS

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[20], 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.

B. SI Model

In this model, the status of peers in a P2P network classified into two classes. One is susceptible, the other infected. Susceptible peers are not sharing any infected files, but are at risk of downloading infected files. When a peer downloads an

infected file, it becomes infected at once. Upon execution, a total of *c* infected files reside in the peer's shared folder. The state progress for all peers in the model is $S \to I$.

In a P2P network with infected files, when a susceptible peer downloads a file, an infected file can be downloaded. It is easy to deduce that the probability of downloading an infected file is proportional to the proportion of infected files in the network. The total number of files in the network is $M(t)+K(t)$, the expected probability of downloading an infected file is $h(t)=\alpha\frac{K(t)}{M(t)+K(t)}$, where α is a adjusting parameter. The constant α reflects the fact that the probability is close related to worm prevalence.

In a time unit, a susceptible peer downloads λ_d files, while the probability of infected files downloaded is $h(t)$, so the probability of a susceptible peer becoming infected is $\lambda_d h(t)$. Therefore, the overall rate of change of S is $-\lambda_d h(t)S(t)$. It is evident that the changing rate of I is the negative of the changing rate of I. When a susceptible peer is infected, the number of infected files increases by c. The rate of change of K is $c\lambda_d h(t)S(t)$. Therefore, the differential equations of the SI model are as follows.

$$\frac{dS(t)}{dt} = -\lambda_d h(t) S(t) \qquad (1)$$

$$\frac{dI(t)}{dt} = \lambda_d h(t) S(t) \qquad (2)$$

$$\frac{dK(t)}{dt} = \lambda_d h(t) S(t) c \qquad (3)$$

$$\frac{dM(t)}{dt} = \lambda_d N(1 - h(t)) \qquad (4)$$

where $N(t) = S(t) + I(t)$

C. SIS Model

In this model, the state progress of a peer can be $S \to I \to S$. When all infected files of an infected peer are naturally death or deleted, the peer returns to the susceptible state. Let λ_{is} denote the average rate, in peers in a time unit, at which infected peers return to the susceptible state. And then, in a time unit there are $\lambda_{is} S(t)$ infected peers to become susceptible. According to analysis in the SI model, the differential equations of the SIS model are:

$$\frac{dS(t)}{dt} = -\lambda_d h(t)S(t) + \lambda_{is} I(t) \qquad (5)$$

$$\frac{dI(t)}{dt} = \lambda_d h(t)S(t) - \lambda_{is} I(t) \qquad (6)$$

$$\frac{dK(t)}{dt} = c\lambda_d h(t)S(t) - c\lambda_{is} I(t) \qquad (7)$$

$$\frac{dM(t)}{dt} = \lambda_d N(1-h(t)) \qquad (8)$$

where $N(t) = S(t) + I(t)$

D. SIR Model

Like the SIS model, the SIR model is classical epidemic model, too. In this model, peers can only exist in three different states: susceptible, infected, or removed (immunized or dead). The state progress of a peer can be $S \to I \to R$. Unlike the SIS model, some proportion of those infected peers change into the immunized state instead of returning to the susceptible state. When an infected peer is removed, it means that all infected files on the peer are deleted and the peer can be infected no more since then. Assuming susceptible peers and infected peers are removed by the proportion λ_{sr} and λ_{ir} per time unit, respectively. Removes occur at rate $\lambda_{sr} S(t) + \lambda_{ir} I(t)$. At the same time, the infected files decrease at rate $c\lambda_{ir} I(t)$ .In this situation, the differential equations are:

$$\frac{dS(t)}{dt} = -\lambda_d h(t)S(t) - \lambda_{sr} S(t) \qquad (9)$$

$$\frac{dI(t)}{dt} = \lambda_d h(t)S(t) - \lambda_{ir} I(t) \qquad (10)$$

$$\frac{dR(t)}{dt} = \lambda_{sr} S(t) + \lambda_{ir} I(t) \qquad (11)$$

$$\frac{dK(t)}{dt} = c\lambda_d h(t)S(t) - c\lambda_{ir} I(t) \qquad (12)$$

$$\frac{dM(t)}{dt} = \lambda_d N(1 - h(t)) \qquad (13)$$

where $N = S(t) + I(t) + R(t)$

As is addressed above, the three models are suitable for the different stages of worm propagation stages, respectively. At the preliminary stage, only the transition of states is $S \to I$, so the SI model can reflect this case. With progress of worm propagation, some users are aware of worms in their hosts and delete those infected files. In addition, some worms can be naturally dead. In the cases, infected peers return to be susceptible. The model corresponding to this case is the SIS model. As passive worms spread, more and more users realize that there exist worms in the P2P networks and take countermeasures (include patching their systems and updating anti-virus software). In this course, a proportion of peers are immunized. Because the proportion of peers immunized is much more than the one of peers returning to the susceptible state, the proportion of peers returning to be susceptible is ignored. The SIR model just reflects this case.

IV. SIMULATION EXPERIMENTS

A. Simulation Description

In order to verify the validity of these models proposed in this paper, i.e. the analytical predictions on worm propagation based on these models are in accordance to the fact, large scale simulations based on these models are carried out. To compare with simulation values, we use the numerical analysis tool: Matlab Simulink, to work out theory values. To simulate the P2P workload and passive worm propagation, we

implemented a simulation framework driven by realistic popular P2P protocols such as BitTorrent and eDonkey, based on the simulation platform Peersim. The simulator first initializes various components, such as nodes and files. Almost all the nodes are initialized to be susceptible and only quit a few nodes are initialized to be infected. To simplify simulation, the same assumptions are abided by in the simulator.

In the next sections, we study how the passive worms propagate under different situations. For each of the experiments, we ran the simulation twenty times and took the average for the plots. We summarize common simulation parameters in Table 2, and we set the default values of some parameters. All simulations use the default values in table 2 except for the parameter being varied by individual. Those parameters which are not in table 2 have the same default value 0.

TABLE II. THE DEFAULT VALUES OF THE PARAMETERS

S(0)	I(0)	λ_d	λ_{sr}	λ_{ir}	λ_{is}	c
10000	10	0.02	0.001	0.002	0.001	10

B. Simulation Evaluation

Large scale simulations show that the predictions of these models keep consistence with the truth of worm spreading on P2P networks. Because of limited space, only a few simulations are depicted. Figure 1-3 compare the results of simulations based on Peersim with the theory results which are worked out by Matlab Simulink. Figure1-3 all show that the simulation results and the theory results form two approximate curves, which illustrates that these models are reasonable and can be used to predict the passive worm propagation in P2P networks.

Figure 4 compares the results of the three models. The number of infected peers of the SI model increases fastest, while this case is evident. Compared with the SI model, the number of the SIS model has a slower increasing rate and in some time units the number keeps invariable i.e. the infection reaches the steady state. The curve of the SIR model goes up at first, then after reaching the peak prevalence, begins to go down.

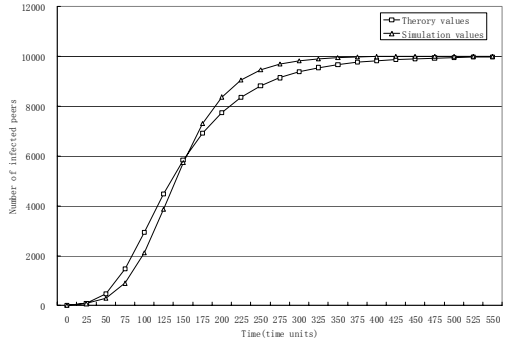

Figure 1. Comparison between the theory values and the simulation values of the SI model

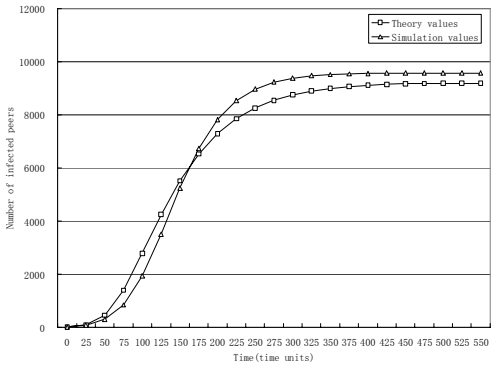

Figure 2. Comparison between the theory values and the simulation values of the SIS model

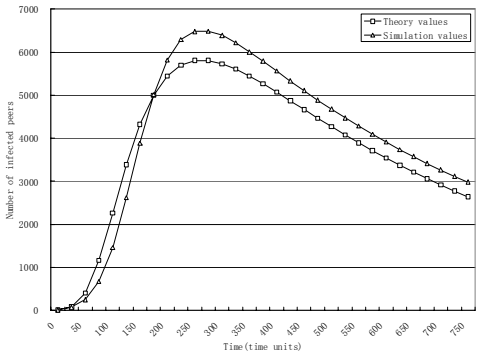

Figure 3. Comparison between the theory values and the simulation values of the SIR model

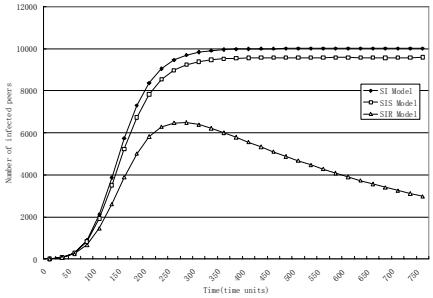

Figure 4. Comparison of numbers of infected peers of the three models

V. CONCLUTIONS

Since the first worm, Morris, arose, worms have been threatening the Internet and other networks. Certainly, the P2P network is no exception. In this paper, we aim at modeling P2P passive worm propagation. Because there are different propagation features at different spreading stages, we obtain three models, which are suitable for different stages. Large scale simulations verify the validity of our models, as means that these models can be used to predict and reflect passive worm expected propagation behaviors. The future work will focus on improving these models to make them to be valid in condition of variable network size, peers adding or leaving, and new files adding.

ACKNOWLEDGMENT

The author would like to thank the anonymous reviewers for their valuable comments and suggestions that improve the presentation of this paper. This work is supported by the National Natural Science Foundation of China under Grant No.60473090 and the joint research project funded by the Royal Society in the UK and by the National Natural Science Foundation of China (NSFC) under Grant No.60711130232.

REFERENCES

[1] S. Staniford, V. Paxson, and N.Weaver. How to Own the Internet in Your Spare Time. In Proceedings of the 11th USENIX Security Symposium, San Francisco, CA, Aug.2002.

[2] "eDonkey2000 server list," http://ocbmaurice.no-ip.org/slist/serverlist.html.

[3] Bittorrent Protocol Specification v1.0,
http://www.bitconjurer.org/BitTorrent/protocol.html

[4] D. Stutzbach, R. Rejaie, and S. Sen. Characterizing unstructured overlay topologies in modern P2P file-sharing systems. In Proceedings of the Fifth ACM Internet Measurement Conference, pages 49–62, Berkeley, CA, Oct.2005.

[5] D. Moore, C. Shannon, and k claffy. Code-Red: a case study on the spread and victims of an Internet worm. In: Proceedings of the Second ACM Internet Measurement Workshop, 2002.

[6] Gnutella protocol development, http://rfc-gnutella.sourceforge.net

[7] Viruslist.com, "P2p-worm.win32.achar.a," http://www.viruslist.com/en/viruses/encyclopedia?virusid=23893, May 2003.

[8] Symantec, "W32.hllw.gotorm," http://securityresponse.symantec.com/avcenter/venc/data/w32.hllw.gotorm.html,August 2003.

[9] Viruscan, "W32/bare.worm," http://www.virus-scan-software.com/latest-virus-software/latest-viruses/w32bare-worm.shtml,2003.

[10] Sophos, "Sophos virus analysis: Troj/krepper-g," http://www.sophos.com/virusinfo/analyses/trojkrepperg.html, July 2004.

[11] C. C. Zou, W. Gong, and D.Towsley, "Code Red Worm Propagation Modeling and Analysis", In Proceedings of9-th ACM Conference on Computer and Communication Security(CCS), Washington DC, November 2002.

[12] Z. Chen, L. Gao, and K. Kwiat, "Modeling the Spread of Active Worms", IEEE INFOCOM, 2003.

[13] C.C. Zou, D. Towsley, W. Gong, and S. Cai, "Routing Worm: a Fast, Selective Attack Worm based on IP Address Information", Univ. Massachusetts Technical Report TRCSE-03-06, November, 2003. http://tennis.ecs.umass.edu/czou/research/routingWormtechreport.pdf.

[14] Wei Yu, "Analyze the Worm-Based Attack in Large Scale P2P Networks", In Proceedings of8th IEEE International Symposium on High Assurance Systems Engineering (HASE'04), 2004.

[15] Wei Yu, "Analyzing the performance of Internet worm attack approaches", In Proceedings of 13th International Conference on Computer ommunications and Networks,2004.

[16] Wei Yu, Corey Boyer, Sriram Chellappan and Dong Xuan, "Peer-to-Peer System-based Active Worm Attacks:Modeling and Analysis", In Proceedings ofIEEE International Conference on Communications (ICC), May 2005.

[17] L. Zhou, L. Zhang, F. McSherry, N. immorlica,M. Costa, and S. Chien. A first look at peer-to-peer worms: Threats and defenses. In Proceedings of the 4thInternationalWorkshop on Peer-to-Peer Systems, Ithaca,NY, Feb. 2005.

[18] M. Ripeanu. Peer-to-peer architecture case study:Gnutella network. In Proceedings of the First International Conference on Peer-to-Peer Computing, Linkoping, Sweden, Aug. 2001.

[19] S. Sen and J.Wang. Analyzing peer-to-peer traffic across large networks. IEEE/ACM Transactions on Networking,12(2), Apr. 2004.

[20] Frauenthal J C. Mathematical Modeling in Epidemiology. Springer-verlag, New York, 1980.