Abstract

P2P worms pose heavy threatens to P2P networks. P2P worms exploit common vulnerabilities in member hosts of a P2P network and spread topologically in the P2P network, a potentially more effective strategy than random scanning for locating victims. Considering that the topology of P2P networks has important effect on P2P active worm spreading, it is very difficult to model propagation of P2P active worms. For this reason, so far few propagation models are proposed. In this paper, we propose a propagation model of active worms in P2P networks based the discrete-time method. The analysis on simulation results shows that there exists some threshold value during spreading of active worms and different infection strategies result in different infection results in P2P networks.

Keywords: Active worms, model, propagation, simulation, P2P networks

1. Introduction

Peer-to-peer (P2P) overlay networks enjoy enormous and ever increasing popularity in real-life deployment. 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]. 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] , the 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. P2P networks provide an ideal venue for new types of worms that prey on common vulnerabilities on the hosts in a P2P network. These worms identify new victims simply by following P2P neighbor information on infected hosts. They are different from the currently popular scanning worms, which probe addresses randomly for new victims, in three important ways. First, they spread much faster, since they do not waste time probing unused IP addresses. Second, they do not generate high rates of failed connections. Finally, they can blend into the normal traffic patterns of the P2P network. The lack of abnormal network behavior makes P2P worms a potentially more deadly threat because most existing defense mechanisms against scanning worms are no longer effective. Because the number of subscribers to a P2P network such as BitTorrent is estimated to be in the millions, P2P worms have the potential to compromise a significant fraction of the Internet population[4].

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 spreading of P2P active worms in P2P networks. In next section, we will introduce active worms in detail. This paper contributes as follows.
1) Propose a propagation model of P2P active worms based on the discrete-time method.
2) Use numerical simulating software, which is developed by us ourselves, to simulate.
3) Refer to the impact of varying initial infection strategies on worm spreading.

The rest of this paper is organized as follows. We simply introduce the existing studies of worm propagation in Section 2. In Section 3, we present a propagation model of active worms. In Section 4, we analyze the properties of worm spreading with simulating experiments. Finally we conclude in Sections 5.

2. Related work

2.1. P2P Worm Classification
According to scanning strategies, Worms can be classified into two classes. one is called as scanning worms, and the other non-scanning worms. Many notorious Internet worms employ a random scanning strategy to find the potential victims. P2P worms tend to use neighbor list to find the potential victims instead of scanning, so P2P worms are non-scanning. According to different attacking ways, we identify three types of non-scanning worms that could leverage P2P networks: passive worms that hide themselves in malicious files and trick users into downloading and opening them; reactive worms that only propagate with legitimate network activities; and active worms that automatically connect to and infect known peers using topological information [5]. Note that the reactive and active worms are analogous to contagion and topological worms.

2.2. P2P Networks

A P2P networks is a group of Internet nodes that construct their own special-purpose networks on top of the Internet. Such a system performs application level routing on top of IP routing. There are two types of P2P networks: structured P2P and unstructured P2P networks. The structured P2P networks are systems in which nodes organize themselves in an orderly fashion, while unstructured P2P networks are ones in which nodes organize themselves randomly. Structured P2P networks boast an efficient lookup mechanism by means of DHTs (Distributed Hash Tables). In the structured P2P system, all P2P nodes maintain the same topology degree, which defines the number of neighbors for each P2P node. For example, one node in d-dimensional CAN maintains 2d neighbors [6]. Conversely, unstructured P2P networks use mostly broadcast search, like the BitTorrent system [7]. In this system, the topology degree is a variable for each P2P node. In this paper, we use BitTorrent system to represent the generalized unstructured P2P networks.

2.3. Existing Modeling Work on Worms

Modeling and analysis of the propagation of worms have been studied for several years. Stanford et al. used the classical simple epidemic model to model the spread of Code Red worm [1,8]. They also concluded that P2P system is well suitable for contagion worm propagation, but they didn't give detailed modeling and analysis. Zou et al. presented two-factor worm model that considered human countermeasures and network congestion effect [9]. Chen et al. presented discrete-time version worm model that considered patching and cleaning effect [10]. Yu et al. researched propagation models of Peer-to-Peer system-based worms in Internet[11,12,13]. Existing work on P2P networks shows that the P2P topologies approximately power-law distributions[14,15].

3. P2P active worm propagation model

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. 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 workload model, to study the active P2P worms. To model in the discrete-time method, it is necessary to explain these parameters and assumptions employed in the following model.

3.1. Model Parameters and Assumptions

The intent of our modeling is to predict the propagation behaviors of a worm which spreads through a P2P network. We make the simplifying assumption as follows.

1) The topology of P2P networks is unchanged, including the number of peers and the degree distribution. This assumption is reasonable for the reason that P2P active worms spread very fast.

2) It takes a time unit that a peer transits from a state to another.

Table 1. Parameters in models

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Descriptive</th>
</tr>
</thead>
<tbody>
<tr>
<td>S(t)</td>
<td>Proportion of susceptible peers at time unit t. Here, S(0)=1.</td>
</tr>
<tr>
<td>I(t)</td>
<td>Proportion of infected peers at time unit t.</td>
</tr>
<tr>
<td>S_k(t)</td>
<td>Proportion of susceptible peers among all peers(susceptible or infected) with degree k at time unit t. Here, S_k(t)+I_k(t)=1</td>
</tr>
<tr>
<td>I_k(t)</td>
<td>Proportion of infected peers among all peers(susceptible or infected) with degree k at time unit t.</td>
</tr>
<tr>
<td>β</td>
<td>Average infection probability, at which a susceptible peer is infected by an infected neighbor.</td>
</tr>
<tr>
<td>γ</td>
<td>Average recovery probability, at which an infected peer returns to be susceptible.</td>
</tr>
<tr>
<td>k</td>
<td>Average degree of a P2P networks</td>
</tr>
<tr>
<td>P(k)</td>
<td>Proportion of peers with degree k among all peers.</td>
</tr>
</tbody>
</table>
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.

### 3.2. SIS Model

To address the effect of the topology of P2P networks in epidemic spreading we shall use the SIS epidemiological model. In this model, peers can only exist in two discrete states, namely, susceptible and infected. These states completely neglect the details of the infection mechanism within each individual. The P2P active worm transmission is also described in an effective way. At each time unit, each susceptible node is infected with probability $\beta$ if it is connected to one infected nodes. At the same time, infected nodes are cured and become again susceptible with probability $\gamma$. Peers run stochastically through the cycle susceptible $\rightarrow$ infected $\rightarrow$ susceptible, hence the name of the model. The SIS model does not take into account the possibility of peer removal due to death or acquired immunization.

**Lemma 1** In a P2P network with initial infected nodes, the probability that any susceptible node is infected is

$$p = \beta k \Theta(I(t)),$$

where $\Theta(I(t)) = \sum_{k} \frac{kp(k)I_{k}(t)}{sP(s)} = \tilde{k}^{-1} \sum_{k} kp(k)I_{k}(t)$

Proof: for any node with degree $k$, because the ratio of the number of infected neighbor nodes to the one of all neighbor nodes is $I_{k}(t)$, there are $kI_{k}(t)$ nodes infected in all its neighbors. Thus, the average number of infected neighbors in the network is $\sum_{k} kp(k)I_{k}(t)$.

Because the average number of neighbor nodes is $\tilde{k} = \sum_{s}sP(s)$, the average infected proportion of nodes is $\ddot{k} = \sum_{k} kp(k)I_{k}(t)$, which is denoted as $\Theta(I(t))$.

Because a node with degree $k$ has $k\Theta(I(t))$ infected neighbors and the probability of a susceptible node being infected by one of its infected neighbors $\beta$, therefore, the probability that the node is infected at time $t$ is

$$p = \beta k \Theta(I(t)).$$

**Theorem 1** In a time unit, the rate at which the proportion of infected node among all nodes with degree $k$ changes is

$$I_{k}(t+1) - I_{k}(t) = \beta k (1 - I_{k}(t))\Theta(I(t)) - \Psi_{k}(t)$$

Proof: According to Lemma 1, a susceptible node with degree $k$ is infected by the probability of $\beta k \Theta(I(t))$. And at time unit $t$, the proportion of susceptible nodes among all nodes with degree $k$ is $(1 - I_{k}(t))$. Hence, there are $\beta k (1 - I_{k}(t))\Theta(I(t))$ susceptible nodes to become infected in a time unit. Meanwhile, there are $\Psi_{k}(t)$ infected nodes returning to be susceptible. Therefore, the change rate of the proportion of infected nodes with degree $k$ in a time unit is

$$I_{k}(t+1) - I_{k}(t) = \beta k (1 - I_{k}(t))\Theta(I(t)) - \Psi_{k}(t)$$

**Theorem 2** In a time unit, the change rate of the proportion of infected nodes in a P2P network is

$$I(t+1) - I(t) = \beta \tilde{k} (1 - \Theta(I(t)))\Theta(I(t)) - \gamma I(t)$$

Proof: $I_{k}(t+1) - I_{k}(t) = \beta k (1 - I_{k}(t))\Theta(I(t)) - \Psi_{k}(t)$

$$I_{k}(t+1) = \beta k (1 - I_{k}(t))\Theta(I(t)) + (1 - \gamma)I_{k}(t)$$

$$p(k)I_{k}(t+1) = \beta k p(k)\Theta(I(t)) + (1 - \gamma)\sum_{k} p(k)I_{k}(t)$$

$$I(t+1) - I(t) = \beta \tilde{k} (1 - \Theta(I(t)))\Theta(I(t)) - \gamma I(t)$$

$$I(t+1) - I(t) = \beta \tilde{k} (1 - \Theta(I(t)))\Theta(I(t)) - \gamma I(t)$$

### 4. Simulation experiments

#### 4.1. Simulation Description

In order to analyze and study on the propagation properties of P2P active worms, we implemented a simulation system based on the simulation platform Peersim. Simulating involves two steps. First, we use the Brite, a popular topology generator, to generate a power-law topology. Second, we run the simulation software with the topology to get simulation results. The simulator first initializes various components, such as infected probability and initial infected peers. Almost all the peers are initialized to be susceptible and only quit a few nodes are initialized to be infected. To clearly compare theory values with simulation values, we put them on the same plots. To
simplify simulation, the same assumptions are abided by in the simulator.

### 4.2. Simulation Evaluation

According to Figure 1-3, no matter what the value of \( \delta = \beta / \gamma \) (called as valid infection rate) is, when it reaches or surpass some value (the threshold), the epidemic spreading of P2P active worms will reach an endemic state with stationary and constant value of the prevalence of infected peers. Moreover, when it reached or surpass some larger value, all peers will be infected (Figure 1). On the contrary, when it is smaller than the threshold, all peers will be susceptible in some time (Figure 3).

Figure 4 examines the effect of initial infected peers with varying degrees on the propagation of active worms. Intuitively, for the initial infected peers with larger degrees, the active worm would spread faster and reach the steady state with larger stationary and constant value of prevalence of infected peers. Figure 4, however, shows that for initial infected peers with varying degrees, the values of prevalence are same. Indeed, initial infected peers with larger degrees will lead to faster spreading of active worms, which suggests that the propagation of active can be throttled by keeping those peers with larger degrees from worms.

Figure 4: Comparison of infectious prevalence with different infection strategies. Selected infection (Max) refer to the case of initial infected peers with 10 maximum degrees, while Selected infection (Min) with 10 minimum degrees.

### 5. Conclusions

In this paper, we aim at modeling P2P active worm propagation in the discrete-time method. Large scale simulations are done in P2P networks with scale-free topology. According to the simulation results, we analyze on the spreading properties of active worms in P2P networks. The analysis on simulation results shows that there exists some threshold value during spreading of active worms. When the valid propagation rate reaches or surpass the threshold value, the epidemic spreading of P2P active worms will reach an endemic state with constant value of the prevalence of infected peers. Or else the infection dies out fast. The simulation results also shows that different infection strategies result in different infection results, which is helpful for us to propose the methods of throttling spreading of worms.

**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. This work is also supported by the important project of Sichuan Normal University of China under Grant No.07ZD018.

References