Peer-to-Peer System-based Active Worm Attacks: Modeling and Analysis

Wei Yu, Corey Boyer[+], Sriram Chellappan[+], Dong Xuan[+]

Computer Science Department,
Texas A&M University, College Station, TX 77843
Email: {weiyu@cs.tamu.edu}, Tel: 972-335-8782
[+]Department of Computer Science and Engineering,
The Ohio State University, Columbus, OH 43210
Email: [+]{boyerp, chellapp, xuan}@cse.ohio-state.edu}, Tel: 614-292-2958

Abstract—Recent active worm propagation events show that active worms can spread in an automated fashion and flood the Internet in a very short period of time. Due to the recent surge of Peer-to-Peer (P2P) systems with large numbers of users, P2P systems can be a potential vehicle for the active worms to achieve fast worm propagation in the Internet. In this paper, we address the issue of the impacts of active worm propagation on top of P2P systems. In particular: 1) we define a P2P system based active worm attack model and study two attack strategies (an off-line and on-line strategy) under the defined model; 2) we develop an analytical approach to analyze the propagation of active worm under the defined attack model and conduct an extensive study to the impacts of P2P system parameters, such as size, topology degree, and the structured/unstructured properties on active worm propagation. Based on numerical results, we observe that a P2P-based attack can significantly worsen attack effects (improve the attack performance) and we observe that the speed of worm propagation is very sensitive to P2P system parameters. We believe that our work can provide important guidelines in design and control of P2P systems as well as active worm defense.

Keywords—P2P System, Active Worm Attacks

I. INTRODUCTION

In this paper, we analyze the impact of Peer-to-Peer (P2P) systems on active worm propagation in the Internet. The propagation of active worms in the Internet enables one to control thousands of hosts by launching distributed denial of service attacks, accessing confidential information, and destroying valuable data. Due to the recent surge of many popular P2P systems with a large number of users, P2P systems can be a potential vehicle for the active worm attacker to achieve fast propagation. It is important to place our work in context.

Active worms have been persistent security threats on the Internet, especially during the last few years. In 2001, the well-known Internet worm Code-Red caused 360,000 hosts to be infected in 10 hours and more than $1.2 billion in economic damage in the first 10 days [1].

P2P computing is becoming an active area for Internet-scale resource sharing and cooperation. The recent surge of P2P applications can be observed by following statistical data collected on Nov. 10, 2003: there are a total of 3,467,860 users in the *FastTrack* P2P system and 103,466 users in the *Gnutella* P2P system [2]. These numbers are still increasing. Recent worm attacks like the *MyDoom* worm have spread themselves over the *Kazaa* P2P system through the P2P file sharing [3]. Due to the recent surge of many popular P2P systems with a large number of users, P2P systems can be a potential vehicle for the active worm attacker to achieve fast propagation. We expect P2P-based worm attacks to be one of the best facilitators of Internet worm propagation and achieve fast worm propagation than existing worm attack approaches, i.e, Code-Red, due to the following reasons: 1) compromising P2P systems with a large number of registered active hosts can easily accelerate Internet worm propagation, as hosts in P2P systems are real and active; 2) some hosts in P2P systems may have vulnerable network and system environments, e.g., home networks; 3) as hosts in P2P systems maintain a certain number of neighbors for routing purposes, worm infected hosts in the P2P system can easily propagate the worm to its neighbors, which continue the worm propagation to other hosts and so on.

In P2P systems, much work has focused on routing efficiency and scalability [4] and some areas of security such as secure routing [5] and *DOS* (denial of service) attacks in the *Gnutella* P2P system [6]. In the active worm area, much work has been done in worm modeling such as the computer virus model [7], active worm spreading model [8], and Code-Red worm model [1]. Besides these, some work has been done on active worm defense system such as containment-based approach to slow down the worm propagation by controlling the worm scan rate [9]. Several attack techniques for effective worm propagation over different types of network systems including P2P systems are discussed in [10], but no detailed modeling and analysis, particularly on the attack propagation over P2P systems, is presented. To the best of our knowledge, there has been no formal/analytical approach devoted exclusively to study different P2P-based attack strategies and analyze the impact of P2P systems on global Internet worm propagation.

The goal of our work is to develop an analytical methodology that can be used qualitatively to better understand the impacts of Internet and active worm propagation that use P2P systems as a vehicle. The highlights of this paper are:

1) We define the P2P-based attack model and study two P2P-based attack strategies: an offline P2P-based hit-list attack strategy and an online P2P-based attack strategy.

2) We develop an analytical approach to analyze the impact of attack strategies and P2P-related factors such as P2P system size, vulnerability of P2P systems, P2P topology degree, and whether the systems are structured or unstructured. We obtain interesting results: P2P-based attacks can significantly worsen the attack effects; both P2P size and P2P topology degree have worsening impacts on the worm attack effects; the unstructured P2P system also has a worsening impact on the worm attack effects.

We believe that results of our work can provide important guidelines for P2P system design and control to address the

concerns of active worm propagation. The rest of paper is organized as follows: some background related to active worms and P2P systems are given in Section II. In Section III, we propose the P2P-based worm attack model. In Section IV, we give the formal analysis results for all attack strategies. In Section V, numerical analysis results and discussions are given. Conclusion of this paper and future work are discussed in Section VI.

II. BACKGROUND

In this section, we will give some background about active worms and P2P systems.

A. Active Worm Attacks

The most famous active worm is the *Morris* worm, which quickly crippled a substantial portion of the Internet in 1988. 'Active worm' is defined in U.S. v Morris: in the colorful argot of computers, a 'worm' is a program that travels from one computer to another but does not attach itself to the operating system of the computer or the computer it infects. It differs from a virus, which is also a migrating program, but one that attaches itself to the operating system of any computer it enters and can infect any other computer that uses the files from the infected computer. In other words, the worm is a software component that, under its own means is capable of infecting one computer system and using it in an automated fashion, to infect another system. This cycle is then repeated and the population of worm-infected hosts grows exponentially, especially in a network environment.

B. P2P Systems

A P2P networked system 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 systems: structured P2P and unstructured P2P systems. The structured P2P systems, such as *CAN* [11], *Chord* [12], *Pastry* [13] and *Tapestry* [14], are systems in which nodes organize themselves in an orderly fashion, while unstructured P2P systems are ones in which nodes organize themselves randomly. Structured P2P systems boast an efficient lookup mechanism by means of *DHT*s (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 [11]. Contrarily, unstructured *P2P* systems use mostly broadcast search, like *Freenet* and *Gnutella* systems [15] [16]. In this system, the topology degree is a variable for each P2P node. In this paper, we use *CAN* to represent the generalized structured P2P systems and use *Freenet* and *Gnutella* systems to represent the generalized unstructured P2P systems.

III. MODELING P2P-BASED ACTIVE WORM ATTACKS

In this paper, we consider three attack models. The first is a purely random-based attack, which is the fundamental attack strategy adopted by many worms. Then, we study two P2P-based attack strategies: an offline P2P-based approach and an online P2P-based approach. For simplification purposes, we do not consider the cooperation of worm-infected hosts to share the attack information. In this sense, the victims could be attacked by different worm-infected hosts at multiple times during the attack runtime, due to non-cooperation of infected hosts.

A. Worm Attack Strategies

1) Pure Random-based Scan (PRS): In this strategy, worm-infected hosts do not have any prior vulnerability knowledge or active/ inactive information of other hosts. The worm host randomly selects the IP addresses of victim targets from the global IP address space and launches the worm attack. When the new host is infected, it continuously attacks the system by using the same methodology. In this paper, this attack strategy is treated as the baseline attack for comparison purposes, as it has been widely adopted by many worms such as, Code-Red-I and Slammer [1][17].

2) Offline P2P-based Hit-list Scan (OPHLS): In this strategy, we assume that worm-infected hosts collect all IP address information of the P2P system offline, denoted as the hit-list. Worm-infected hosts launch the attack against hosts in the hit-list. In this attack strategy, all newly infected hosts continuously attack the hit-list until all hosts in the hit-list have been scanned. Then, all worm-infected hosts continue to attack the system via PRS.

3) Online P2P-based Scan (OPS): In this strategy, after joining the P2P system at the system's initial time, the worm-infected host immediately initiates an attack against its P2P neighbors with its full attack capacity. At the same time, the worm-infected hosts can also attack the system via PRS if extra attack capacity is available. To illustrate this, an example is given: Say A_1 is the worm infected host with attack capability 5 (i.e., it is able to attack 5 hosts simultaneously) and A_1 has three P2P neighbors B_1, B_2, and B_3. It starts to use 60% of its attack capability to attack B_1, B_2, and B_3 and the rest of the attack capability (40%) to attack the system via PRS. Assuming that B_2 and B_3 are vulnerable hosts and infected, these two newly infected hosts will continuously attack their P2P neighbors and the system by repeating the attack cycle of A_1. After that, A_1 will use 100% of its attack capability to attack the system via PRS. The detailed algorithm is as follows:

Algorithm 1 – OPS
(P2P node i as the worm-infected host with attack capability -S)

1. *Finds m P2P neighbors, i.e., $G = \{h_1, h_2, \ldots, h_m\}$*
2. *While (G is not empty)*
   If ($S \geq m$)
      Scan m P2P neighbors
      Use the S-m scan capability to scan the system via PRS
      G = Null
   else
      Scan S neighbors, i.e., $h_1, h_2, \ldots, h_s$
      $G = G - \{h_1, h_2, \ldots, h_s\}$
3. *Attack the system via PRS*

B. Model Parameters

In order to formally analyze P2P-based attack strategies, we list the following most important parameters, which will have an impact on worm attack effects.

1) Attacker parameters: Attack scan rate *S* and the system's initial infected worm instances M_0 are two of the most important parameters from the worm attacker perspective. Intuitively, the larger the values, the faster is the propagation.

2) P2P system parameters: For P2P-based systems, the following parameters need to be considered: i) Topology degree in Structured P2P systems: The topology degree defines the number of P2P neighbors maintained by the P2P host locally. Based on our analysis, this parameter only has an impact on the online P2P-based worm attack strategy. For the structured P2P system, the topology degree is actually a constant. ii) Topology degree in Unstructured P2P systems: The topology degree of unstructured P2P systems can be modeled by the mean value associated with the topology degrees for all P2P hosts, as the topology degree for each P2P host is actually a variable. A nontrivial development related to complex networks discovered that for most large networks, including the Internet, metabolic, protein networks, social networks and email systems, the distribution of the host topology degree follows the power-law distribution. Based on previous studies [16][18], we model the topology of the unstructured P2P systems using the power law distribution. In power law theory, the spread in the number of edges of the diverse network hosts is characterized by the degree distribution *P(k)* which gives the probability that a randomly selected host has exactly *k* edges. We consider the distribution:

$$P(k) = C_1 \frac{\varpi}{k^{\sigma}} \quad (1),$$

where ϖ is the mean value of topology degree, C_1 is constant for each given ϖ, and $\sigma \in [1,8]$ is the parameter used to represent the power law degree. iii) Size of P2P system: This parameter defines the number of hosts in the P2P system. It will have an impact on both offline and online P2P-based attack strategies. iv) Vulnerability of P2P systems: This parameter measures the vulnerability for P2P hosts. As we mentioned, the host in the P2P can be used in less protected environments, such as a home environment. Table 1 lists all parameters and notations in this paper.

C. Assumptions

We assume that the system IP address space is the IP address space of IPv4, or 2^{32}. In the IPv4 address space, some valid IP addresses are not actively utilized, are non-routable, or are even not applicable to the host (based on the previous statistical result [19], only 24% of available addresses are used by active hosts).

We assume that there are two logical systems: one is called a 'super-P2P' system, which generalizes P2P systems in the Internet, and the other is called 'non-P2P' system, which represents the rest of system. In both 'super-P2P' system and 'non-P2P' system, we assume that a number of hosts are vulnerable. As our analysis considers the average case, we assume that each host in 'super-P2P' or 'non-P2P' system has a certain probability to be vulnerable.

In this paper, we do not consider the time taken for the infected host to find the vulnerability of victims and assume that the worm infecting one victim takes unit time. At the system's initial time, we assume that there are a certain number of infected hosts and infected hosts are already in the 'super-P2P' system.

T	Total IP addresses in the system
S	Scan rate of worm infection host (number of victims being able to be scanned simultaneously)
M_0	Initial worm hosts at the system's initial time
P_1	Probability of the IP address being utilized by the host
P_2	Probability of real hosts in the system being vulnerable
R_1	Size of 'super-P2P' system
θ	Topology degree of structured P2P system
σ	Power law degree for the unstructured P2P system
r_j	Topology degree of host j for the unstructured P2P system with power law distribution
ϖ	Mean value of topology degree for the unstructured P2P system with power law distribution
P_3	Probability of hosts in the P2P system to be vulnerable (P_3:P_2 defines the comparative vulnerability between the 'super-P2P' system and 'non-P2P' system)
M(i)	Number of infected hosts at time i in the whole system (*M(0)* is the number of initial infected hosts in the system)
N(i)	Number of vulnerable hosts at the time i in the whole system (*N(0)* is the number of vulnerable hosts to be infected at the system's initial time)
E(i)	Number of newly infected hosts added at step I in the whole system (initial value $E(0) = 0$)
M(i,X)	Number of infected hosts at step i, where *M(i, 0)* is the infected hosts in the 'non-P2P' system and *M(i, 1)* is the infected hosts in the 'super-P2P' system (*M(0, 0)* is the initial infected hosts in 'non-P2P' system and *M(0, 1)* is the initial infected hosts in the 'super-P2P' system)
N(i,X)	Number of vulnerable hosts at step i, where *N(i, 0)* is the vulnerable hosts in 'non-P2P' system and *N(i, 1)* is the vulnerable host in 'super-P2P' system (*N(0, 0)* = $T*P_1*P_2$ is the initial vulnerable hosts in 'non-P2P' system and *N(0, 1)* = R_1*P_3 is the initial vulnerable hosts in 'super-P2P' system)
E(i,X)	The newly infected hosts added at step i, where $E(i,0)$ is the newly infected hosts added at step i in 'non-P2P' system and $E(i,1)$ is the newly infected hosts added at step i in 'super-P2P' system

Table 1: Notations in this paper

IV. WORM ATTACK ANALYSIS

To better understand the characteristics of the active worm spread, we adopt the epidemic dynamic model for disease propagation. In order to make it flexible for analyzing P2P-based attack strategies, we use discrete time to conduct recursive analysis and approximate the worm propagation [7] [8]. In the following, we list several theorems which give out the formulas to compute *E(i)*, *M(i)* and *N(i)* under different P2P based active worm attack strategies. The proof of these theorems can be found in Appendix A.

Theorem 1: For the *PRS* approach, with $M(i)$ and $N(i)$ at time i, the next tick will have

$$E(i+1) = ((N(i) - M(i))[1 - (1 - \frac{1}{T})^{SM(i)}] \quad (2)$$

newly infected hosts, where *E(0) = 0, N(0) = T*P1*P2, M(0)=M0.* Also following recursive formulas are used to calculate *M(i+1)* and *N(i+1)*.

$$M(i+1) = M(i) + E(i+1), \quad (3)$$
$$N(i+1) = N(i) - E(i+1).$$

This theorem is similar to Theorem 1 in [8], both of which are for *PRS*. The differences are in initial values such as *N(0)* and notations.

Theorem 2: For the *OPHLS* approach, with $M(i,1)$ and $N(i,1)$ at time *i* in the 'super-P2P' system, the next tick will have

$$E(i+1,1)=\begin{cases}(N(i,1)-M(i,1))[1-(1-\frac{1}{R_1})^{SM(i,1)}] & if \quad M(i,1)<R_1P_3 \\ 0 & if \quad M(i,1)\geq R_1P_3\end{cases} \quad (4)$$

With $M(i,0)$ and $N(i,0)$ at the time *i* in the 'non-P2P' system, the next tick will have

$$E(i+1,0)=\begin{cases}0 & if \quad M(i,1)<R_1P_3 \\ (N(i,0)-M(i,0))[1-(1-\frac{1}{T})^{SM(i,0)+SM(k,1)}] & if \quad M(i,1)\geq R_1P_3\end{cases} \quad (5)$$

where $M(0,1)=M_0$, $M(0,0)=0$, $N(0,1)=R_1P_3$, $N(k,0)=T*P_1*P_2-R_1*P_3$, $M(k,0)=M(k-1,1)$ ($k=\min(i)$ for i satisfying the condition $M(i,1)\geq R_1P_3$), $M(i)=M(i,1)+M(i,0)$, $N(i)=N(i,1)+N(i,0)$.

Also there are the following recursive formulas.

$$M(i+1,0)=M(i,0)+E(i+1,0), \quad (6)$$
$$N(i+1,0)=N(i,0)-E(i+1,0),$$
$$M(i+1,1)=M(i,1)+E(i+1,1), \quad (7)$$
$$N(i+1,1)=N(i,1)-E(i+1,1).$$

In order to differentiate between the structured and unstructured P2P systems in the *OPS* attack strategy, we use *OPSS* (Online P2P-based scan for Structured P2P system Strategy) and *OPUS* (Online P2P-based scan for Unstructured P2P system Strategy) to represent the structured and unstructured online P2P-based attacks respectively. We have obtained the following theorems.

Theorem 3: In *OPSS* approach, with $M(i,0)$ and $N(i,0)$ at time *i* in the 'non-P2P' system, the next tick will have

$$E(i+1,0)=(N(i,0)-M(i,0)[1-(1-\frac{1}{T})^{(M(i,0)+M(i,1))*S-\min\{\theta,S\}*E(i,1)}]. \quad (8)$$

With $M(i,1)$ and $N(i,1)$ at the time *i* in the 'super-P2P' system, the next tick will have

$$E(i+1,1)=(N(i,1)-M(i,1)[1-(1-\frac{1}{R_1})^{\min(\theta,S)E(i,1)}], \quad (9)$$

where $M(0,0)=0$, $M(0,1)=M_0$, $N(0,0)=T*P_1*P_2-R_1*P_3$, $N(0,1)=R_1*P_3$, $N(i)=N(i,1)+N(i,0)$. Also we have the same recursive formulas here as in *Theorem 2*.

Theorem 4: In *OPUS* approach, with $M(i,0)$ and $N(i,0)$ on average at time *i* in the 'non-P2P' system, the next tick will have

$$E(i+1,0)=(N(i,0)-M(i,0)[1-(1-\frac{1}{T})^{(M(i,0)+M(i,1))*S-\sum_{j=1}^{E(i,1)}\min(r_j,S)}]. \quad (10)$$

With $M(i,1)$ and $N(i,1)$ on average at the time *i* in the 'super-P2P' system, the next tick will have

$$E(i+1,1)=(N(i,1)-M(i,1)[1-(1-\frac{1}{R_1})^{\sum_{m=1}^{E(i,1)}\min(r_j,S)}]. \quad (11)$$

where r_j is the topology degree for host *j*, which is one of *E(i,1)* newly infected hosts at time *i*, $M(0,0)=0$, $M(0,1)=M_0$, $N(0,0)=T*P_1*P_2$, $N(0,1)=R_1*P_3$, $M(i)=M(i,1)+M(i,0)$, $N(i)=N(i,1)+N(i,0)$. Also we have same recursive formulas as *Theorem 2*.

V. NUMERICAL RESULTS AND DISCUSSIONS

A. Simulation Model

1) Metrics: For all scan strategies, the system attack performance is defined as follows: the time taken *t* (X axis) to achieve successful infection ratio – *infected host number/ total vulnerable host number* (Y axis). The higher the performance value, the worse is the attack effect. *2) Evaluation Systems:* The general system is defined by the tuple: <*A, T, S, M_0, P_1, P_2, P_3, R_1, θ, σ, ϖ*>, representing the system configuration parameters. *A* determines the attack strategy and can be one of <*PRS, OPHLS, OPSS, OPUS*>. Other parameters have the same definition in Table 1. As we are only focusing on selected important parameters that are sensitive to the P2P-based attack strategies, the following parameters are set with constant values (*$T=2^{32}$, S=6, $M_0=1$, $P_1=0.25$*) in all our simulations. *3) Evaluation Method:* We use numerical analysis to obtain performance data.

B. Performance Result

In this section, we report the performance results along with observations. Due to the space limitations, we only present a limited number of cases here. However, we found that the conclusions we draw here generally hold for many other cases we have evaluated. All the data are shown start at time 45, as the infection ratio is very small (close to 0) in the time interval [0, 45] due to the large number of vulnerable hosts in the simulation. Recall that our definition of Infection Ratio is *infected host number/ total vulnerable host number*.

1) Comparison among all attack strategies

Fig. 1 shows the data on the sensitivity of attack performance to different attack strategies. The general system is configured as <*, *2^{32}*, 6, 1, 0.25, 0.2, 0.2, 10000, 4, 4, 3*>. From this figure, we make the following observations: i) The P2P-based attack strategies overall outperforms the *PRS* attack strategy. For example, in the worm fast propagation phase (linear increase – from simulation time 50 to 70), the P2P-based approach can achieve 50%-100% performance increase over the *PRS* attack strategy. The result matches our expectation: attacking the P2P system achieves a higher successful scan rate, which can significantly improve the attack performance. From the defense perspective, the P2P-based attack will be a very challenging issue. ii) *OPHLS* attack strategy achieves the best performance compared to all other online-based attack strategies. The reason is that without the comparatively slow run-time search preceding the attack for the online P2P-based attack strategy, *OPHLS* has all P2P system information at the beginning of attack, which makes the worm propagation fast. As we do not consider the time taken for the offline collection of P2P system information, this strategy can be considered an upper bound for the P2P-based attack.

2) The Sensitivity of P2P System Size

Fig. 2 shows the data on the sensitivity of attack performance under different P2P system sizes. The general system is configured as <*, 2^{32}, 6, 1, 0.25, 0.2, 0.2, *, 4, 4, 3>. In this figure, the P2P size $R_1 \in [1000,10000]$ and in the legend, *A(#)* implies that the attack uses approach *A* with # representing the P2P size. We observe the following: with the P2P size increase, the attack performance becomes consistently better for all attack strategies. The result matches our expectations: the larger is the size of the P2P system, the higher is the scan hit probability achieved, as all the hosts in the P2P system are active hosts.

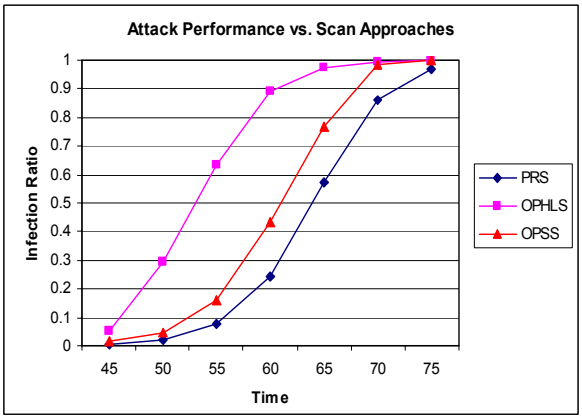

Figure 1: Performance Comparison of All Attack Strategies

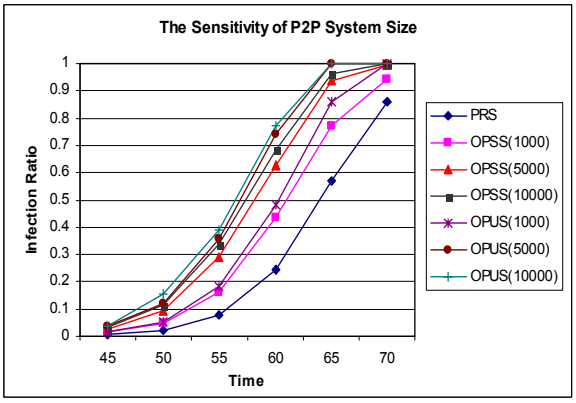

Figure 2: The Attack Performance Sensitivity to P2P System Size

3) The Sensitivity of P2P Topology Degree

Fig. 3 shows the data on sensitivity of attack performance for different P2P topology degrees. The general system is configured as <*, 2^{32}, 6, 1, 0.25, 0.2, 0.2, 10000, *, *, 3>. In this figure, *OPSS (#)* defines the P2P-based attack for the structured P2P system with # representing the topology degree. *OPUS* (#,*) defines the P2P-based attack for the unstructured P2P with power law parameters: # representing ϖ and * representing σ (set as 3). We have made the following observations: for the structure-based P2P attack strategy, an increase in topology degree achieves better attack performance. This matches our expectation - larger topology degrees make more P2P hosts open to the attacker and speeds up the worm propagation.

4) The sensitivity of Unstructured P2P Parameter

Fig. 4 shows the data on the sensitivity of attack performance for the unstructured P2P parameters. The general system is configured as <*, 2^{32}, 6, 1, 0.25, 0.2, 0.2, 10000, 4, 4, *>, and σ is between [2, 8]. From this figure, we note the following observations: i) For the unstructured P2P system with the fixed mean value of topology degree ϖ, a lower power law degree σ achieves better attack performance. The reason can be explained: due to the large tail property of the power law distribution, a smaller value of σ means that the probability of a host having k neighbors (P(k)) is high (from equation 1). Thus, due to the increased connectivity, the infection ratio is higher when σ is smaller. ii) When ϖ is equal to θ (structured P2P topology degree), the unstructured P2P system achieves better attack performance than structured P2P system. The result matches our expectation: large tail property of the power law distribution implies large number of hosts with large topology degree, which increases the worm propagation speed.

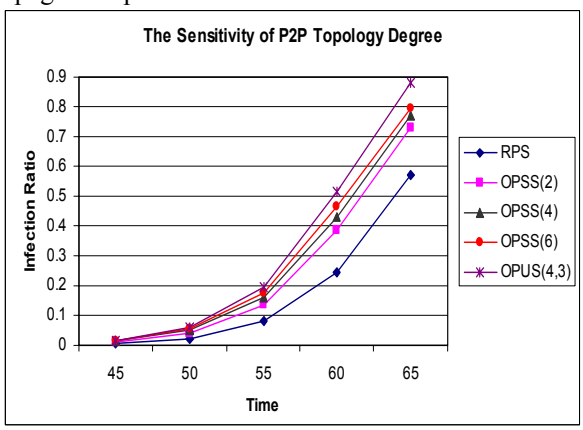

Figure 3: Structure and Unstructured P2P-based Attack Strategies with Different Topology Degrees

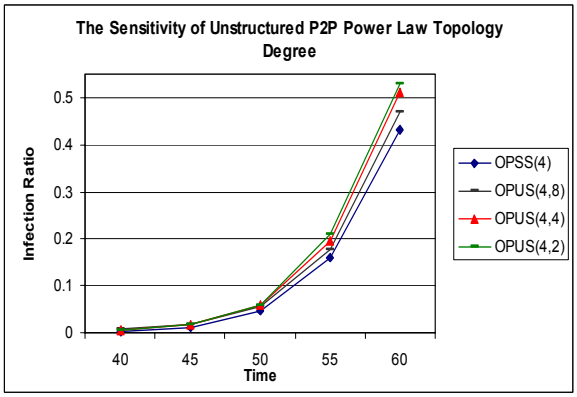

Figure 4: The Sensitivity of Unstructured P2P Parameter

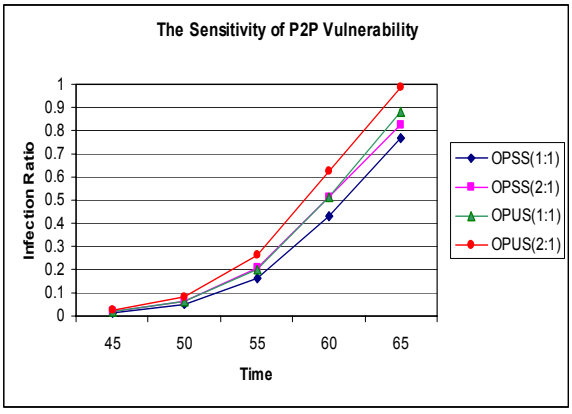

Figure 5: Structured & Unstructured P2P Systems with Different Vulnerabilities

5) The sensitivity of P2P vulnerability

Fig. 5 shows data on the sensitivity of the attack performance for different vulnerabilities in both structured and unstructured P2P systems. The general system is configured as <*, 2^{32}, 6, 1, 0.25, 0.2, *, 10000, 4, 4, 3>. Here P_3:P_2 is selected as 1:1 and 2:1. We note the following observations: With the increase in vulnerability of P2P hosts, better attack performance is achieved consistently for all P2P-based attack strategies. The result matches our expectation: a larger vulnerable value causes more vulnerable hosts to be infected in a given time. More infected hosts added during the attack run-time makes the worm propagation fast.

VI. CONLCUSIONS

In this paper, we have studied the impacts of the propagation of active worms on top of P2P systems. Active worm are quite potent. In a parallel direction, P2P systems are turning to be quite popular. The threats and effects of worm propagation on top of P2P systems will result in significant damages as illustrated by our analysis prompting more attention on active worm attacks in P2P system design and control. While the different P2P parameters and properties such as system size, degree and topology etc. have their own impacts on the P2P system performance (hit ratio, average path length, convergence etc), configuration and selection of such parameters and properties should also consider their potential impacts on active worm propagation, particularly in environments where threats exist without any efficient defense mechanisms.

Practically speaking, the effectiveness of configuring P2P parameters and properties to defend active worms are limited. A proactive defense system needs to be in place. Our current research focuses on this. While a P2P system can be a vehicle for active worm propagation, it can also be a backbone to defend against active worm attacks. The preliminary idea is to build up a self-organized overlay on the top of the original P2P system to perform the worm detection and defense. Such an overlay is composed of P2P hosts which cooperate among each other and are trusted. Effective worm detection schemes and design of a lightweight, adaptive and distributed defense architecture are the key focus for our current investigation in this regard.

ACKNOWLEDGEMENT

Dong Xuan's work is supported in part by National Science Foundation under grant No.ACI-0329155. Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

REFERENCES

[1] C. C. Zou, W. Gong, and D.Towsley, "Code Red Worm Propagation Modeling and Analysis", In *Proceedings of 9-th ACM Conference on Computer and Communication Security (CCS)*, Washington DC, November 2002.
[2] Slyck news, "http://www.slyck.com".
[3] Mydoom, "http//www.f-secure.com/tools".
[4] J. Xu, J., Kumar, A., and X. Yu, "On the Fundamental Tradeoffs between Routing Table Size and Network Diameter in Peer-to-Peer Networks", *IEEE Journal on Selected Areas in Communications*, vol 22, no 1, pp. 151-163, January 2004.
[5] M. Castro, P. Druschel, A. Ganesh, A. Rowstron, and D. S. Wallach, "Secure Routing for Structured Peer-to-Peer Overlay Networks", In *Proceedings of the 5-th USENIX Symposium on Operating Systems Design and Implementation (OSDI)*, Boston, Massachusetts, December 2002.
[6] N. Daswani and H. G. Molina, "Query-Flood Dos Attack in Gnutella", In *Proceedings of the 9-th ACM conference on Computer and Communications Security (CCS)*, Washington, DC, November 2002
[7] J. O. Kephard and S.R. White, "Directed-graph Epidemiological Models of Computer Virus", In *Proceedings of 1991 Computer Society Symposium on Research in Security and Privacy*, Oakland, CA, May 1991.
[8] Z. S. Chen, L.X. Gao, and K. Kwiat, "Modeling the Spread of Active Worms", In *Proceedings of IEEE INFOCOM,* San Francisco, CA, March 2003.
[9] S. Staniford, "Containment of Scanning Worms in Enterprise Networks", *Journal of Computer Security*, 2003.
[10] 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, August 2002.
[11] S. Ratnasamy, P. Francis, M. Handley, R. Karp, and S. Shenker, "A Scalable Content Addressable Network", In *Proceedings of ACM SigComm.*, San Diego, 2001.
[12] I. Stoica, R. Morris, D. Karger, M. F. Kaashoek, and H. Balakrishnan, "Chord: A Scalable Peer-to-peer Lookup Service for Internet Applications", In *Proceedings of ACM SIGCOMM 2001*, San Deigo, CA, August 2001.
[13] A. Rowstron and P. Druschel, "Pastry: Scalable, Distributed Object Location and Routing for Large-scale Peer-to-peer Systems", In *Proceeding of IFIP/ACM International Conference on Distributed Systems Platforms (Middleware)*, Heidelberg, Germany, November, 2001
[14] B. Y. Zhao, L. Huang, J. Stribling, S. C. Rhea, A. D. Joseph, and J. Kubiatowicz, "Tapestry: A Resilient Global-scale Overlay for Service Deployment", *IEEE Journal on Selected Areas in Communications*, vol. 22, no. 1, pp. 41-53, January 2004.
[15] L. A. Adamic, R. M. Lukose, A. R. Puniyani, and B. A. Huberman, "Search in Power-law Networks", *Physical Review E*, Vol. 64, 2001.
[16] M. Ripeanu and I. Foster, "Mapping the Gnutella Network: Macroscopic Properties of Large-Scale Peer-to-Peer Systems", In *Proceedings of 1-th International Workshop on Peer-to-Peer Systems (IPTPS)*, Cambridge, MA, March 2002.
[17] D. Moore, V. Paxson, S. Savage, C. Shannon, S. Staniford, and N. Weaver, "Inside the Slammer Worm", *IEEE Magazine of Security and Privacy*, August 2003.
[18] P. Silvey and L. Hurwitz, "Adapting Peer-to-Peer Topologies to Improve System Performance", In *Proceedings of the Hawaii International Conference on System Sciences*, Hawaii, Jan, 2004.
[19] A. Zeitoun and S. Jamin, "Rapid Exploration of Internet Live Address Space Using Optimal Discovery Path", In *Proceedings of IEEE GLOBECOM (Next Generation Networks and Internet)*, San Francisco, CA, December 2003.

Appendix A

In the following, we will give out the proofs of Theorem 1, 2, 3, and 4 discussed in Section IV. As mentioned before, Theorem 1 is similar to one in [8]. The differences are in notations and some initial values such as *N(0)*. Hence, the proof is also similar but with different initial values and notations to the one in [8]. For the sake of completeness and ease of understanding other theorems, we devote space for the proof of Theorem 1 here.

A1. Proof of Theorem 1: As $M(i)$ infected hosts can generate $S*M(i)$ scans in an attempt to infect other hosts (Recall that S is the scan rate of an effected host). We need to prove

$E(i+1 \mid K) = ((N(i) - M(i))[1 - (1 - \frac{1}{T})^{K}]$ for K scans. By induction, when $K = 1$, since there are $N(i)$ vulnerable hosts at step i ($N(0)=T*P_1*P_2$), one scan adds newly infected hosts

$$\frac{N(i)}{T} - \frac{M(i)}{T}$$

, where the first term represents the number of newly infected hosts and the second term considers the factor of infected hosts being scanned and infected at multiple times due to the non-cooperation of worm infected hosts. Assume that we have

$$E((i+1) \mid J) = ((N(i) - M(i))[1 - (1 - \frac{1}{T})^{J}]$$

for J scans. Then, the $J+1$ scan can be divided into two steps: the first J scans and the last scan. For the last scan, there are two possibilities: adding a newly infected host or not. Let variable $Z=1$, if the last scan hits a vulnerable host that has not yet been infected and let $Z=0$ otherwise. Then

$$\begin{aligned} & E((i+1) \mid J+1) \\ & = (E((i+1) \mid J) + 1)P(Z=1) + E((i+1 \mid j)P(Z=0) \\ & = E((i+1) \mid J) + P(Z=1) \\ & = (N(i) - M(i))[1 - (1 - \frac{1}{T})]^{J+1}. \end{aligned}$$

Therefore, we have

$$E((i+1) \mid (J = SM(i)) = (N(i) - M(i))[1 - (1 - \frac{1}{T})]^{SM(i)}.$$

With $E(i+1)$, we can calculate $M(i), N(i)$ at step i.

A2. Proof of Theorem 2: We classify the worm attack into two steps:

i) *Attack hosts in the 'super-P2P' system with offline collected IP addresses (hit-list).* As the size of 'super-P2P' is R_1 and the vulnerability probability of P2P host is P_3, there are $N(0,1) = R_1 P_3$ vulnerable hosts in the 'super-P2P' system. Similarly, there are $N(i,1) - M(i,1)$ vulnerable hosts that have not been infected at step i, one scan adds following newly infected hosts in the 'super-P2P' system at step i:

$$(N_1(i,1) - M(i,1))[1 - (1 - \frac{1}{R_1})^{1}].$$

With the similar analysis approach shown in *Theorem 1*, we have following result for $SM(i,1)$ scans at step i:

$$E(i+1,1) = ((N(i,1) - M(i,1))[1 - (1 - \frac{1}{R_1})^{SM(i,1)}], if \quad M(i,1) < R_1 P_3.$$

When $M(i,1) \geq R_1 P_3$, the 'super-P2P' system has been fully attacked, we have

$$E(i+1,1) = 0 \qquad if \quad M(i,1) \geq R_1 P_3 .$$

ii) *Attack the 'non-P2P' system*: after attacking the 'super-P2P' system, all infected hosts continuously attack the 'non-P2P' system. The initial step k starting to attack 'non-P2P' system can be calculated by following formula: $k = \min(i)$ for i satisfying $M(i,1) \geq R_1 P_3$. As the total vulnerable host number in the 'super-P2P' system is $R_1 P_3$ and the vulnerable host number in the whole system is $T * P_1 * P_2$, $N(k,0) = T * P_1 * P_2 - R_1 * P_3$ represents the vulnerable hosts that have not been scanned at step k. Thus, the total number of scans at step i $(i \geq k)$ is

$$SM(i,0) + SM(k,1)$$

and we have

$$E(i+1,0) = \begin{cases} 0 & if \quad M(i,1) < R_1 P_3 \\ (N(i,0) - M(i,0))[1 - (1 - \frac{1}{T})^{SM(i,0)+SM(k,1)}] & if \quad M(i,1) \geq R_1 P_3. \end{cases}$$

A3. Proof of Theorem 3: For the 'super-P2P' system, we can easily prove the result by following: one scan for each infected host can add $[(N(i,1) - M(i,1)] \frac{1}{R_1}$ newly infected hosts in the 'super-P2P' system. Considering the P2P network with topology degree θ, the worm-infected hosts can maximally generate $\min(\theta, S)$ scans to attack other P2P hosts simultaneously. As a newly added infected hosts is $E(i,1)$ at step i, the total scan number is $\min\{\theta, S\}E(i,1)$. Using the similar induction method, we have

$$E(i+1,1) = (N(i,1) - M(i,1)[1 - (1 - \frac{1}{R_1})^{\min(\theta,S)E(i,1)}].$$

For the 'non-P2P' system, there are total $M(i,0) + M(i,1)$ worm-infected hosts in the whole system and $\min\{\theta, S\}E(i,1)$ scans are applied to attack the 'super-P2P' system at step i. Thus, the total scan number for the 'non-P2P' system at step i is

$$(M(i,0) + M(i,1)) * S - \min(\theta, S) * E(i,1)$$

and we have

$$E(i+1,0) = (N(i,0) - M(i,0)[1 - (1 - \frac{1}{T})^{(M(i,0)+M(i,1))*S - \min\{\theta,S\}*E(i,1)}].$$

A4. Proof of Theorem 4: For the 'super-P2P' system, we can easily prove the result by following: one scan for each infected host can add newly infected hosts

$$[(N(i,1) - M(i,1)] \frac{1}{R_1} .$$

As there are $E(i,1)$ infected hosts attacking the 'super-P2P' system at i step and each infected host has topology degree r_j, each worm-infected host can simultaneously generate $\min(r_j, S)$ scans to attack other P2P hosts and the total scan number is $\sum_{j=1}^{E(i,1)} \min(r_j, S)$. By using the similar induction method, we have

$$E(i+1,1) = (N(i,1) - M(i,1)[1 - (1 - \frac{1}{R_1})^{\sum_{j=1}^{E(i,1)} \min(r_j, S)}].$$

For the 'non-P2P' system, there are total $M(i,0) + M(i,1)$ worm-infected hosts in the whole system and $\sum_{j=1}^{E(i,1)} \min(r_j, S)$ scans are applied to attack the 'super-P2P' system at step i. Thus, the total scan number for the 'non-P2P' system at step i is

$$(M(i,0) + M(i,1)) * S - \sum_{j=1}^{E(i,1)} \min(r_j, S)$$

and we have

$$E(i+1,0) = (N(i,0) - M(i,0)[1 - (1 - \frac{1}{T})^{(M(i,0)+M(i,1))*S - \sum_{j=1}^{E(i,1)} \min(r_j, S)}].$$