Defense against Passive Worms in P2P Networks

Bo Zhan [1], Laurissa Tokarchuk [1], Chaosheng Feng [2], Zhiguang Qin [2]

1. Department of the Electronic Engineering, Queen Mary, University of London, London E1 4NS

2. School of Computer Science &Engineering, University of Electronic Science and Technology of China,
Chengdu Sichuan 610054;

Abstract

In recent years, internet users are getting more and more familiar with P2P networks, which allow them to share their content files with scalability and reliability. However, due to the fast speed of data and file exchange in P2P networks, the hosts in the P2P system are especially vulnerable to P2P passive worms. These worms hide themselves in malicious files and trick users into downloading and opening them. In this paper, we first propose a model of passive worm propagation in unstructured P2P networks. Then we propose and evaluate a patch dissemination method to combat the spread of passive worms in P2P networks. This method mimics worm-like behavior to disseminate the patch files. The simulation results indicate that it has a great contribution to constraining passive worm propagation and reducing the peak number of the infected hosts in the P2P network.

1 Introduction

P2P networks are now being used as the most popular means of data and file sharing. Internet users are familiar with using the P2P network to exchange data, media files, especially some popular resources in the internet, which can be downloaded millions of times with the help of P2P networks. The unstructured P2P file-sharing networks, such as Bit-Torrent and Kazaa [1], are very popular due to its scalability and flexibility. Bit-Torrent, the most familiar P2P application for many Internet users, has more than 10 million users [2]. Another popular P2P file sharing network is the eDonkey2000 network, which alone typically has over 2 million users connected at any given time [3].

For most unstructured P2P file-sharing systems, the more popular the resource is, the faster downloading speed the user can achieve. This means the file and data exchange in P2P networks is much faster than the traditional way of server and client, which give the worms and malicious code an ideal environment in which to propagate. Typically a worm is launched in a P2P network by combining it with some popular resource in the P2P network. As a result, most of the hosts will get infected very fast due to the popularity of the resource file.

Furthermore, some P2P worms have less abnormal network behavior which makes them more difficult to detect. The passive worm is one of the examples. The hosts can only be infected by the user's activation of the passive worm. The actual act of downloading a passive worm file is usually identified as normal behavior and therefore its penetration of the network can be quite high before any alert is generated. It is detected as normal network behavior as neither the user nor the anti-virus software is able to detect the latest passive worms before downloading it.

Passive worms use to propagate slowly in early internet applications due to their inability to autonomously scan and infect other nodes. However, in a P2P network, they can duplicate themselves quickly among the hosts, e.g. the Benjamin worm [4]. This makes the study of passive worm's propagation strategy and techniques to slow down their propagation vital to the continued operation of P2P networks. Since it is difficult to detect the passive worm before downloading it, the most direct way to defend is patching the host as soon as possible. The traditional way of patch dissemination is server and client style, which has been proved to be much slower comparing to the propagation of internet worms [5]. In this paper, we investigate several methods to disseminate the patch by using the P2P network

The rest of the paper is organized as follows: In Section 2, we present the background of passive worms study and introduce some related work in this area. In Section 3, we first describe our modeling of the epidemic spreading of passive worms in unstructured P2P networks, and then we add the P2P patch dissemination modeling into the simulator and run simulations to evaluate the method's performance on restraining the spread of passive worms. The Section 4 is the conclusion and future work.

2 Background

2.1 Passive Worms in P2P networks

Most P2P worms are non-scanning worms which use the neighbor list to find victims instead of using a random scanning strategy which is a normal behavior for most of the scanning worms. There are three types of non-scanning worms: passive worms, reactive worms and proactive worms. The passive worms hide themselves and trick the user to download and execute them. The reactive worms can only propagate with legitimate network activities. The proactive worms automatically connect to and infect the peers using topology information [6]. Some researchers define the passive worms, which attach to files and propagate with user activations, as viruses [7]. We do not make such a distinction for the purpose of this paper.

In this paper we mainly focus on passive worms. Passive worms hide themselves in popular P2P resources by embedding malicious code in executable files. This kind of strategy has historically made passive worms unpopular in P2P networks because most of the files shared in the early P2P network were MP3 files or some other media files, which were not usually executed directly by the user [8]. However, more recent popular P2P systems, like Bit Torrent, Kazza, eDonkey2000 and so on, provide the users much easier access to executable files, which makes passive worms become a more dangerous threat to the safety of the P2P network users.

The passive worms generally behave in the following way: Firstly they embed themselves in the popular executable files in the P2P network and make a few copies in the sharing folder of the infected user. Once another user downloads the files and executes them, the worms duplicate themselves and create a few new copies in the sharing folder, which increases their possibility of being downloaded by the other vulnerable users. Since the user can only be infected after the file is executed, the downloading of the passive worms are, most of time, treated as legitimate P2P network behavior. This makes it quite difficult to detect passive worms.

2.2 Related Work on Internet Worms

Some work has been done to study the internet worm's behavior before. The majority of work has focused on the propagation model of the worms: Thommes and Coates introduced an epidemiological propagation model of P2P virus to study the P2P virus spread in the P2P network [9]. Moore *et al* focused on the case study of Code-Red, a terrible internet worm fast infected lots of computers in 2001 [10]. Zou *et al* also contributed to the study of the Code-Red propagation model [11]. In [12], Staniford *et al* presented the concept of contagion worm, a type of passive worm, and indicated that the P2P network was quite suitable for the passive worm's propagation. They also introduced a classic simple epidemic model to study the spread of Code-Red worm. Although beyond the scope of this study, the propagation model for active worms in P2P networks has also been the subject of recent research [13][14][15].

Furthermore, some researchers focused on the defense against the internet worms: J. Sandin introduced a worm detection method by using P2P network [16]. M. Costa et al. also presented similar idea of monitoring the abnormal behavior with the help of P2P network [17]. Some other worm detection methods are also presented in [18][19][20], but most of them are defense methods against scanning worms which can be detected due to their abnormal network behavior. In [5], Srinivas *et al* introduced a P2P based patch dissemination method to constrain the fast propagation of internet worms. They provided meticulous mathematical analysis but not detailed simulation result.

3 Simulations and Analysis

3.1 P2P Passive Worm Propagation Model

In this section we introduce a simplified propagation model to study the passive worms in unstructured P2P network, including Kazaa, eDonkey2000 and Gnutella [21]. We are using the simplified model because P2P networks are too complex to use an analytical approach to model the worm propagations. We make the simplifying assumption as follows [22]:

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 downloading, 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.

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.

N(t)	Number of all hosts on the P2P network at time unit t, here it is a constant. N(0)=10000.
S(t)	Number of susceptible hosts at time unit t. S(0)=9950.
I(t)	Number of infected hosts at time unit t. S(0)=50.
R(t)	Number of immunized host at time unit t. In SI model, R(0)=0.
K(t)	Number of infected files at time unit t. K(0)=500.
M(t)	Number of uninfected files at time unit t. M(0)=47300.
h(t)	Possibility of downloading an infected file at time unit t. $h(t) = \dfrac{K(t)}{M(t)+K(t)}$.
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. $\lambda_d = 0.02$.
c	When an infected file is downloaded and executed, *c* infected files are generated in the file-sharing folder. *c*=10.
i	Time Intervals to download the Patch File
a	No. of neighbors to probe when a peer decide to download /spread the Patch File

Table 1: Parameters in SI Model and Patch Dissemination

In our former work of simulating the propagation model of passive worm in unstructured P2P network, we have introduced 4 passive worm propagation models: SI model, SIS model, SIR model and SIRE model. For SI model, the node can only be susceptible or infected and once become infected, there is no possibility for the node to get recovered. In SIS, SIR and SIRE model, there is possibility for infected node to get back into susceptible or even recovered status. More detailed description about these models can be found in [22].

In SI model, the status of peers can be classified into two classes. One class 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 $S \rightarrow I$.

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 l 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 $h(t) = \dfrac{K(t)}{M(t)+K(t)}$.

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

3.2 Patch Dissemination against Passive Worms

Based on the passive worm propagation model above, we introduce several patch dissemination methods, which spread the patch files by using the P2P network to slow down the propagation speed of the passive worms in unstructured P2P network. We only use the SI model so we can clearly observe the effect of the P2P patch dissemination. Once the user downloads the patch file, it will change into immune status and cannot be infected again.

For each simulation result in this section, we ran the simulation 20 times and took the average for the plots. Table 1 summarizes the common simulation parameters used.

Patch Mode 1

For Patch Mode 1, we assume the patch files are downloaded as a normal P2P file. This means the patch file has to compete with the other popular P2P files to get the P2P network resources. In the following simulation, we set R (0) =500, which means there are 500 patch files available at the beginning of the simulation.

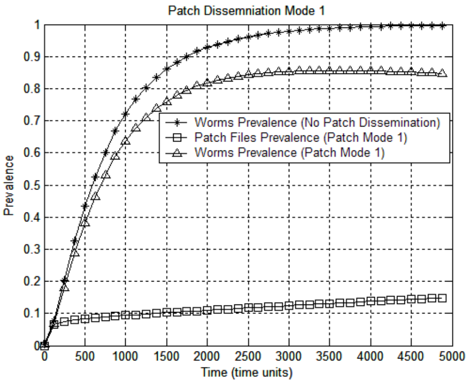

Figure 1: Worm prevalence with Patch Mode 1

As expected, Figure 1 shows that the patch file’s prevalence grows very slowly compared to the passive worms’. This is because passive worm can duplicate itself into multiple copies. This increases the probability for passive worm to be downloaded. So the patch file cannot spread as fast as the worms. Because of the low speed of patching the nodes, Patch Mode 1 is obviously not a suitable approach to throttle down the passive worm fast propagation.

Patch Mode 2

In Patch Mode 2, we suggest a separated P2P patch file sharing system in the P2P network. This means the anti-virus software or firewall just uses the same P2P network to update intermittently by downloading the patch file. In this assumption, each user should try to download the patch file or update its anti-virus software or firewall before it tries to download a share file in the P2P network. We set R (0) =10 in the following simulation. This means there are 10 nodes have the patch file at the very beginning of the simulation.

One concern of Patch Mode 2 is the interference from the patch dissemination. In fact, the patch dissemination consumes some of the P2P network resource, especially when the node probes its neighbors for the latest patch file. In order to reduce the interference from the downloading of the patch file, we should increase the time intervals for each node to update its anti-virus software or firewall.

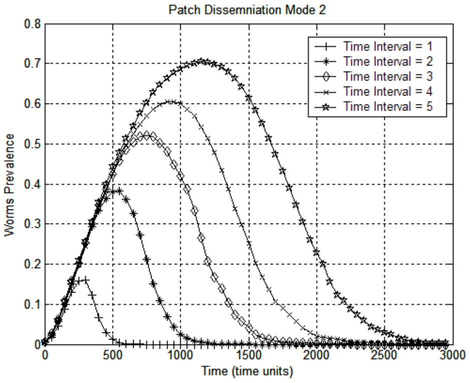

Figure 2: Worm prevalence with Patch Mode 2

As Figure 2 shows, with longer time intervals, we can reduce the interference from the patch dissemination. However, the cost of longer time intervals is more users get infected and the patch file dissemination get slower.

Patch Mode 3

To balance the performance and interference of patch dissemination, we increase the maximum number of neighbors to probe when a node decides to download a patch file. This kind of strategy makes the patch dissemination look more like a worm behavior because each node’s random probing its neighbors for the latest patch file. In the following simulation we set the time interval to 20 time units and change the parameter of maximum probe attempts.

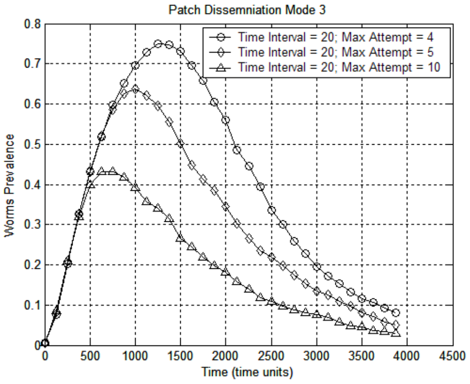

Figure 3: Worm prevalence with Patch Mode 3

As Figure 3 shows, more attempts to download the patch file can provide better performance for the patch file dissemination. This is because more probes to node’s neighbors increase the probability of the unrecovered node to get the patch file. Patch Mode 3 can at least reduce the interference of the patch dissemination in the time domain though it actually consumes similar network resource compare to Patch Mode 2.

Patch Mode 4

For Patch Mode 4, the suspicious or infected node would not probe its N neighbors to download the patch file. However, the recovered node (node already has the patch file) scans its N neighbors and sends the patch file to those who do not have the patch file. To avoid wasting too much P2P network resources sending patch files to nodes that already have them, we introduces a parameter of ineffectual attempt proportion: *p*. At each time unit, a recovered node would only continue to scan and send the patch file to its next neighbor when the proportion of ineffectual scanned neighbors, i.e. nodes which already have the patch file, are less than *p*. In the following simulation, *p* equals 50%.

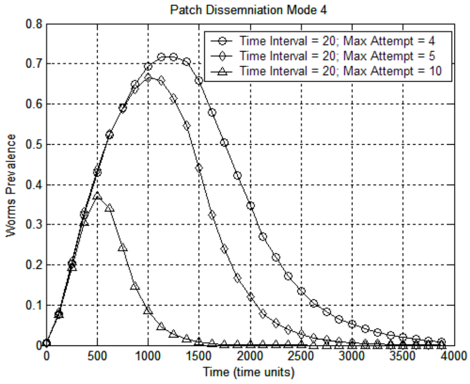

Figure 4: Worm prevalence with Patch Mode 4

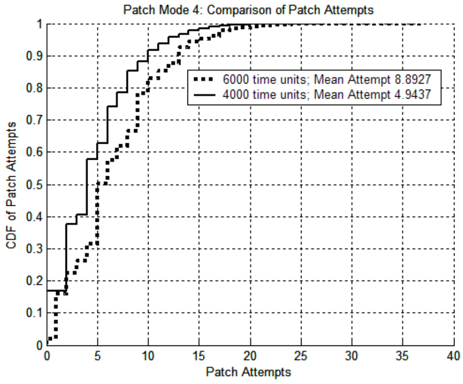

Figure 5: CDF of Nodes' Patch Attempts for Patch Mode 4

As Figure 4 shows, sending the patch file to more neighbors, as illustrated by comparing the difference between Max Attempt = 4 and Max Attempt = 10, provides better performance. However, Patch Mode 4 wastes some network resources to scan the neighbors already have the patch file. For the simulation of Figure 5, we set the max attempts to spread the patch file to 4 and the figure shows the cumulative distribution frequency (CDF) of attempts for each node to spread the patch file. For example, there are 50% nodes make less than 5 attempts to spread the patch file for the 6000 time unit simulation, while about 60% in the 4000 time unit simulation. This figure shows that the longer the simulation goes on, the more resources are wasted because of ineffectual scanning. The introduction of a decay factor on *p* would help mitigate the unnecessary waste of network resources shown in Figure 5.

Patch Mode 5

Patch Mode 5 combines Patch Mode 3 and Patch Mode 4. The unrecovered (infected and susceptible) node probes N neighbors to download the patch file and the recovered node sends the patch file to its N randomly selected neighbors as well. This strategy should spread the patch file much faster and so constrain the propagation of passive worms.

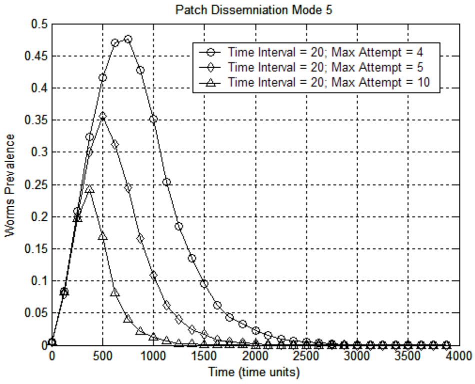

Figure 6: Worm prevalence with Patch Mode 5

As Figure 6 shows, more attempts to download and spread the patch file can provide better performance for Patch Mode 5. With maximum attempt 4, Patch Mode 5 can constrain the peak proportion of infected nodes to 47%. However, Patch Mode 5 consumes more network resource than Patch Mode 3 and 4 because of the probe and scan to spread the patch file to more nodes.

Comparison of Patch Mode 3, 4 and 5

In the following simulations, the general parameters for each Patch Mode are the same.

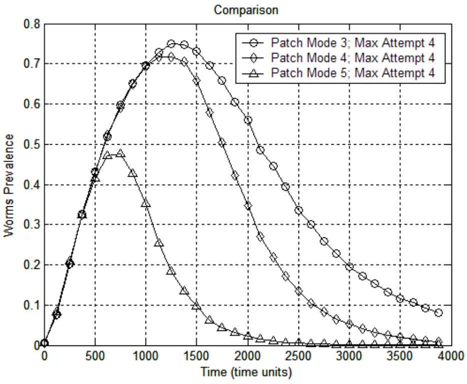

Figure 7: Comparison of Worm prevalence

From Figure 7 we can see that Patch Mode 4 has a little better performance than Patch Mode 3 in spreading the patch file and constraining passive worm's propagation. And apparently, Patch Mode 5 provides the best performance.

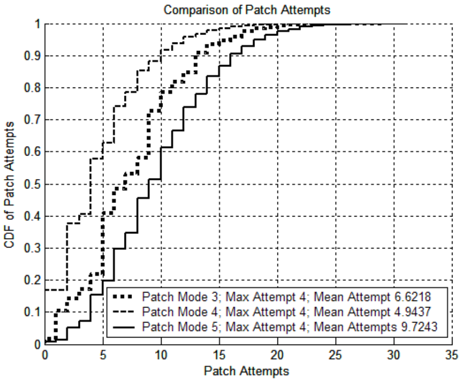

Figure 8: CDF of Nodes' Patch Attempts for Patch Mode 3 4 5

As Figure 8 shows, Patch Mode 4 consumes least network resource. However, as discussed before, Patch Mode 4 would consume more network resources if the simulation runs longer because of the ineffectual scan. And Patch Mode 5 provides the best performance with the cost of the most network resource consumed. The introduction of a decay factor on ineffectual attempt proportion *p* should also help to reduce the unnecessary waste of network resource for Patch Mode 5.

4 Conclusions and Future Work

In this paper, we first introduced a passive worm's propagation model and then use patch dissemination method to fight against the passive worm's propagation in the unstructured P2P network. The simulation proves the effectiveness of the P2P patch dissemination. However, in order to control the interference from the patch dissemination, we increase the time intervals of patch file downloading, which slows down the patch file spread speed. To balance the performance we increase the number of neighbors to probe/scan when an unrecovered node decides to download a patch file or a recovered node tries to spread the patch file to its neighbors. The simulation results show better performance with more probe/scan attempts. Since the passive worm's downloading is difficult to be detected, our work proves that using P2P network to spread the patch files is an effective way to constrain the passive worm's fast propagation.

The future work mainly concerns improving the passive worm's propagation models for more complicated network condition and modeling the other two classes of P2P worms. Another issue is studying users download behavior by monitoring the users' download history. Once the user's download footprint is recorded, it can then be analyzed by using artificial intelligence technology to predict the user's next download behavior. This kind of information can be used to locate the vulnerable file downloader so improve the patch file dissemination's efficiency or improve the performance of other security methods against internet worms.

References

[1] KaZaA Homepage, http://www.kazaa.com

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

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

[4] M. Singer. Benjamin worm plagues KaZaA. Internetnews.com, May 2002

[5] Srinivas Shakkottai , R. Srikant, *"Peer to peer networks for defense against internet worms"* in Proceedings from the 2006 workshop on Interdisciplinary systems approach in performance evaluation and design of computer & communications sytems, Article No. 5

[6] Guanling Chen, Robert S. Gray.*Simulating non-scanning worms on peer-to-peer networks*. In Proceedings of the 1st international conference on Scalable information systems, Hong Kong, China, 2006

[7] N.Weaver, V. Paxson, S. Staniford, and R. Cunningham. "*A taxonomy of computer worms*" in *Proceedings of the1st ACM workshop on Rapid Malcode*, Washington, DC,Oct. 2003.

[8] F-Secure, "F-secure hoax information pages: Mp3virus,"http://www.f-secure.com/hoaxes/mp3.shtml, 1998.

[9] Thommes R, Coates M. *Epidemiological modeling of peer-to-peer viruses and pollution[C]*. In: The 25th Annual IEEE Conference on Computer Communications. IEEE Press, Barcelona, Spain, 2006. 15-26.

[10] D. Moore, C. Shannon, and J. Brown, *"Code-Red: a case study on the spread and victims of an Internet worm,"* in Proceedings of Internet Measurement Workshop (IMW), Marseille, France, November 2002.

[11] C. C. Zou, W. Gong, and D. Towsley, *"Code Red Worm Propagation Modeling and Analysis,"* in 9th ACM Conference on Computer and Communication Security (CCS'02), Washington DC, USA, November 2002.

[12] S. Staniford, V. Paxson, and N. Weaver, *"How to 0wn the Internet in Your Spare Time,"* in Proceedings of the 11th USENIX Security Symposium (Security '02), San Francisco, CA, USA, August 2002.

[13] 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.

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

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

[16] J. Sandin, *"P2P systems for worm detection,"* in DIMACS Workshop on large scale attacks, Piscataway, NJ, USA, September 2003.

[17] M. Costa, J. Crowcroft, M. Castro, A. Rowstron, L. Zhou, L. Zhang, and P. Barham, *"Vigilante: End-to-End Containment of Internet Worms,"* in Proceedings of the 20th ACM Symposium on Operating Systems Principles (SOSP '05), Brighton, United Kingdom, October 2005.

[18] Xia Chunhe ( ), Shi Yunping, Li Xiaojian, GAO Wei, *"P2P worm detection based on application identification"* in Front. Comput. Sci. China 2007, 1(1): 114−122

[19] Geetha Ramachandran, Delbert Hart, *"A P2P Intrusion Detection System based on Mobile Agents"* in ACM Southeast Regional Conference, Proceedings of the 42nd annual Southeast regional conference, Pages: 185 - 190

[20] Guofei Gu, Monirul Sharif, Xinzhou Qin, David Dagon, Wenke Lee and George Riley, *"Worm Detection, Early Warning and Response Based on Local Victim Information"* in Proceedings of the 20th Annual Computer Security Applications Conference, Pages: 136 - 145

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

[22] Zhiguang Qin, Chaosheng Feng, Laurence Cuthbet, Laurissa Tokarchuk, *"Propagation Models of Passive Worms in P2P Networks"*