International Journal of Security and Its Applications
Vol.9, No.11 (2015), pp.271-280
http://dx.doi.org/10.14257/ijsia.2015.9.11.25

A Trust-based Immune Strategy for File Pollution in P2P Networks

Xianwei Xu

Department of Information Technology, Nanjing Forest Police College, Nanjing, Jiangsu, China, 210023
xxw_1025@sohu.com

Abstract

Focusing on the file pollution problem exists in unstructured P2P file sharing systems, we propose a strategy named TrustIs by combining the trust mechanism and the immune strategy of complex network to resist the propagation of polluted files. With the help of peer reputation, trusted immune peers and backup peers can be screened out. Also, the files and indices can be identified based on their object reputation and index reputation so as to provide reference for immune pees to filter out the polluted files or indices. Meanwhile, an immune token ring network is set up for monitoring peers' behavior and implementing the refreshment of the immune resources. The results of the simulations show that TrustIs can effectively resist the propagation of polluted files.

Keywords*: P2P, file pollution, immune strategy, trust mechanism*

1. Introduction

The peer degrees of unstructured P2P network approximately follow the power-law distribution and have the characteristic of scale-free network which characteristic can make the network robust to the random breakdown and vulnerable to the malicious attack [1]. Computer virus is a critical problem on Internet security which has been studied for many years. Its transmissibility, hidden and destructiveness are great threats to our Internet security. Meanwhile, the openness and dynamics of P2P networks make file pollution cause problems such as wasting of resources and inefficiency to the network. P2P networks like KazaA, FastTrack and Overnet are all suffering the problem that more than 50% files are polluted [2]. Since computer virus has the same dynamics property with file pollution, methods that used for combatting virus are also available for file pollution.

Immune strategy is widely used in controlling the propagation of polluted files, especially in the medical field. Vaccination for some special diseases to get people immunity and produce antibodies to combat the virus is considered as an effective application. Work in [3] proposed an object reputation mechanism based on the artificial immunity theory. For the purpose of reducing the possibility of sharing polluted files by malicious peers, the authors combines the artificial immunity with the feedback similarity to choose suited neighbor peers. In another work [4], an immune strategy based on the combination of tracking file propagation chain and immune control was proposed. By peer alarm, tracking propagation chain, immune response and antibody memory to control file pollution. In another aspect, the trust mechanism was also well used to resist malicious attack, such as Scrubber [5] which adopted the peer reputation, Credence [6] which adopted the object reputation and Hybrid [7] which adopted both the former two reputations. The advantages are that we can have an accurate judgment on the peer or file before transaction so that we can filter out the malicious ones without downloading. Thus,

ISSN: 1738-9976 IJSIA

the trust mechanism can not only avoid the unnecessary load resource consumption, but also resist the malicious attack effectively.

Abstracting the specific immune process to find out an effective immune strategy focusing on the propagation process in complex network will be important since a reasonable immune strategy will play an important role in controlling the propagation of polluted files. However, the traditional immune strategies are facing defects such as large resource consumption or low efficiency during the selection of immune peers. These defects make the strategies not applicable for large scale and distributed P2P network. Therefore, a new immune strategy TrustIs based on trust mechanism is proposed in this paper, hoping to resist the propagation of polluted files by selecting the immune peers and immune resource peers based on reputation. Our main work is the following:

1) Combining the anti-virus methods used in human society and computer network to get an immune strategy which has high efficiency and low resource consumption for the resistance of file pollution in P2P network.

2) Using the peer reputation to get credible immune peer, also, using object reputation and index reputation to get authentic files and indices.

3) An immune sub-network is set up to make communication between immune peers and refreshment of immune resources more efficient and accurate.

2. Related Works

Many immune strategies have been used to resist virus propagation in complex network field. Random immune [8] method selects some immune peers in a totally random pattern which means the opportunity for high degree peers and low degree peers to be selected are equal. However, this method has been proven that it needs immunize almost all the peers in order to achieve the goal which is of course not suitable for reality. Target immune [9] method only selects few peers with high degree as the immune peer based on the uneven feature of peer degree in scale-free network. But this method needs to know all the peers' degree as prerequisite which is also not suitable for reality. Acquaintance immune [10] method avoids the problem of target immune which needs to know all the peers' degree and it selects p peers randomly. Then the immune peers come from these p peers' one or more neighbor peers. Although this method costs less resource because of the random selection, it has less recognition degree which makes the strategy less efficient. Also, when facing large amount of malicious peers, it is unreliable for random selection. Work in [11] designed a message delivery algorithm that peer who detects the virus attack will deliver the virus message to its neighbor peer who has the highest degree, and then the neighbor peer will continue the delivery until the virus message is known among the sequence of all the peers with highest degree. The results of experiments show that in the Gnutella network with 10000 peers and its power-law exponent γ=2.0, the virus message needs 36 steps to be delivered around the network. But P2P network is disassortative network in which high degree peers always connect with low degree peers. So it is too idealistic to get highest degree peer in every step. Moreover, as the scale of Gnutella gets larger, relying on one single path among millions of peers is inefficient. All the above methods are some classic immune strategies based on peer degree, paper [12] used peer betweenness to find peers with high transfer ability to be the immune peer. Betweenness is the number of shortest paths which go through a peer or an edge in the network. It can reflect the network load that a peer or an edge can bear in a specific network topology, *i.e.* the transit communication capability. The betweenness-based strategy uses Floyd algorithm to calculate all the shortest paths to count the number of paths which go through every node or edge and the experimental results show that this strategy is better than other degree-based strategies in efficiency and resource consumption. DeepCure [13] proposed a strategy based on target immune which takes

peer degree, betweenness and availability into consideration when selecting immune peers. The authors select p peers and then these peers' neighbor and neighbor's neighbor peers according to the limitations as the immune peers. In this way, only partial information of network topology needs to be known and the immune resource consumption and efficiency can also be taken care of. But in the environment of file pollution, the malicious peers cannot be neglected. So the reputation as well as the management of immune resource is very necessary.

3. Trust Mechanism

3.1. Calculation of Peer Reputation

Every peer's reputation is composed of two parts: direct trust and indirect trust. $f_n(j)$ is the nth feedback after peer i transacted with peer j and we set 1 to represent peer i's satisfaction with peer j's file while -1 represents dissatisfaction. Formula (1) is the calculation of peer i's direct trust to peer j.

$$D_{ij} = \begin{cases} \dfrac{\sum_{n=1}^{m} f_n(j)}{m}, & \mathrm{m} \neq 0 \\ 0, m = 0 \end{cases} \tag{1}$$

Where m is the number of transactions made between peer i and j.

Indirect trust $R(j)$ can be calculated with based on the feedbacks made by the peers who have transacted with peer j. We call those peers who have transacted with peer j feedback peers and their peer reputation will reflect the feedback credibility which means feedbacks made recently are more credible than the older ones. Then we can get peer j's indirect trust $R(j)$ by the following formulae.

$$f_i(j) = \begin{cases} f_i(j), f_i(j) > 0 \\ \beta * f_i(j), f_i(j) < 0 \end{cases} \tag{2}$$

$$R_n(j) = \frac{\sum_{i \in N} NR_F * f_i(j)}{NR_F} \tag{3}$$

$$R(j) = \frac{\sum_{n=0}^{N-1} g_n * R_n(j)}{\sum_{n=0}^{N-1} g_n} \tag{4}$$

Where β is a punishment factor to make sure that the decrease degree of reputation caused by a malicious behavior must be larger than the increase degree caused by a normal behavior. NR_F is the feedback credibility. $f_i(j)$ is feedback made by the ith peer who has transacted with peer j. N is the peer set in which all the peers have transacted with peer j. g_n is the attenuation function to make older feedback less credibility. Then the peer reputation will be the weighted summation of direct trust and indirect trust as shown in Formula (5).

$$NR = \alpha D_{ij} + (1\text{-}\alpha)\ R(j) \tag{5}$$

Where α is a weighted factor. Every peer can choose the value of α according to their own willing. A higher α means the peer is more likely to trust himself while a lower α means the peer trusts the other peers' experiences more.

3.2. Calculation of Object Reputation and Index Reputation

Object reputation can be calculated based on the weighted summation of the feedbacks Vote(O) on the object and the feedback peer's reputation NR. The feedbacks on the object also get two values: 1 and -1.

$$OR = \frac{\sum_{i=0}^{N-1} NR_i * Vote_i(O)}{\sum_{i=0}^{N-1} NR_i} \tag{6}$$

$$IR = \frac{\sum_{j=0}^{N-1} NR_j * Vote_j(I)}{\sum_{j=0}^{N-1} NR_j} \tag{7}$$

Where OR and IR are object reputation and index reputation respectively. The former trust models never take index reputation into consideration, so the file pollution is very serious in some unstructured P2P networks like eDonkey [14] and FastTrack [3] in which file resources are searched by indices and the feedback peers' reputation. If the value of object reputation or index reputation is higher than 0, the object or index is considered to be authentic.

4. TrustIs Immune Strategy

4.1. Selection of Immune Peer

Previous immune strategy like random immune, target immune and acquaintance immune have some defects about selecting immune peers such as large consumption of resources, demands of all peers' information around the network, low efficiency for selecting peers randomly. These defects make the strategies not suitable for complex and dynamic P2P network. Moreover, because of the distributed feature of P2P and there exists too many versions of polluted files, it is not possible to have a central server to maintain a database providing information about polluted files. We propose the immune strategy which combines the peer reputation to select the immune peer. In this way, we can avoid selecting malicious peers and no need for all the peers' information.

Immune peers play important role in immune strategy since they are the key nodes (Hub) in the network and they should isolate polluted files to prevent them from spreading. According to the analysis of peer importance in complex network, the most important peer in the network should has high degree and betweenness [15]. Therefore, immune peers should be the peer with high reputation (to make sure it is credible), high degree (to make sure the propagation paths can be reduced after immunizing the peer), high betweenness (to make sure the peer has high transit communication ability). Here's our method: we first select p peers randomly, then to every peer in the p peers, we select top two peers in the peer's neighbor and neighbor's neighbor according to the above limitations. So one of the top two peers will be the immune peer (IN), the other one will be backup peer (BN). They will monitor each other's reputation and behavior, also they will share the immune resource list.

The reason that we choose immune peer in the two-hop range of the p peers is that the P2P network has degree-degree correlation which is an important factor that affects the immune strategy. The degree-degree correlation is usually defined as the distribution of a specific peer's nearest neighbor peer's average degree. Technical networks like Internet and P2P are both degree-degree correlative networks, also known as disassortative complex networks [16]. Peers with high degree in these networks are more likely to connect with low degree peers, and the file pollution will be propagated along the path of high-degree peer → low-degree peer → high-degree peer. So we choose to immunize the peers in the two-hop range of a peer to control the propagation effectively.

Pseudo code of immune peer's selection:

(1) Select p nodes randomly, say, ；
(2) WHILE(immune peers are not enough)
(3) FOR(i=0；i<P；i++)
(4) In the two-hop range of choose top two peers with highest reputation, degree and betweenness. The two peers will be the immune peer IN and backup peer BN respectively;
(5) END FOR;
(6) Execute the immune resource insertion algorithm；
(7) END WHILE;
(8) Set up the immune sub-network with immune peers and share the immune resources;

4.2. Selection and Insertion of Immune Resource

Immune resource is important to the immune effect and it should be a list of file identifier, *i.e.* immune resource list (iTable), providing reference for immune peers to distinguish polluted files. Traditional C/S structure is very convenient for immune resources to be inserted into the network since the server only needs to send the newest defend command or virus definition to the immune peers. So the previous immune strategy did not mention too much about immune resource. However, in the distributed P2P network, immune resources are all disperse which may cause huge workload consumption for immune resource peers to send the file identifications in the flooding way, let alone the resources' accuracy. Here, we still use reputation, degree and betweenness to select resource peers in the p peers' neighbor peers and we select top K peers to be the immune resource peers (RN). The iTable that resource peers provide contains the object reputation or index reputation and the corresponding index. Every immune peer converges all the iTables owned by its surrounded resource peers as its immune resource list and shares it with the backup peer. If the immune resource exists the duplication of one object, then we use the object's average reputation. While the immune peer finds some object's reputation is lower than 0, then it will delete the file or index. If the object's reputation is 0, then the immune peer waits for the refreshment of the reputation. The immune resource list will be maintained by LRU algorithm and it will be refreshed every week. Here is the format of iTable:

Table 1. *iTable*

Hash of object	Reputation	Hash of index
Vhashi	*ORi/IRi*	*Ihash1* *Ihash2* *…* *Ihashn*

Pseudo code of immune resource's insertion:

(1) For every immune peer $IN_i, 0 \le i < P$;

(2) FOR(i=0; i<P; i++)

Select top K peers with high reputation, degree and betweenness as the resource peers;

(3) RN records the reputation of every version in its iTable;

All the RNs around INi send their iTables which will be converged by INi who will use the information to refresh its own iTable;

(4) END FOR;

4.3. Immune Sub-Network

4.3.1. Topology of the Sub-Network: After the immune peers are selected, every immune peer and its backup peer stores a neighbor list which records the addresses of the precursor and successor immune peers as well as their backup peers. For the purpose of every immune peer can share its iTable and to avoid these peers sending messages without order which will lead to the information redundancy and high load consumption. Moreover, in case the malicious peers interfere the resource share, we make the immune peers to form a token ring network logically. Token is a data frame which can be held by one peer at one time. Peer who holds the token should send it to the next immune peer as well as its iTable. After that, the immune peer also needs to send a feedback to its last immune peer to acknowledge it the token has been sent out. The next immune peer who receives the token will refresh its immune resource list first and then send out the token and its iTable as its precursor does.

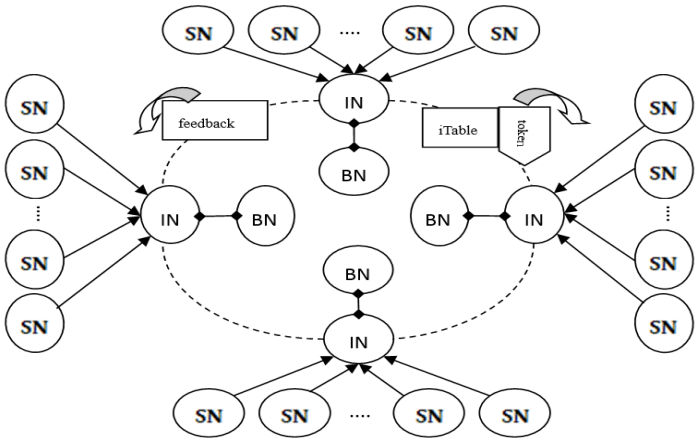

Figure 1. Immune Sub-Network

4.3.2. Immune peer dropping out the network: Peers who drop out the network will be classified into two kinds: normal and abnormal. While the abnormal dropping out may have two situations: peer holds the token and peer with no token. We will discuss the situations in the following.

(1) Normal dropping out. Before the peer drops out, it should send the iTable and token (if it has) to its backup peer. Then the backup peer will be the immune peer and its new backup peer will be selected according to the limitations. The new immune peer and backup peer will inform other peers.

(2) Abnormal dropping with token. When the immune peer's precursor wants to send its iTable, it needs to set up the TCP connection which needs the peers to be online. If the precursor detects its successor is offline, it will reproduce a new token and send its iTable

and the new token to the successor's backup peer. The backup peer who receives the token will inform other peers that it has replaced the former immune peer.

(3) Abnormal dropping without token. Also, the precursor detects that its successor is offline; it sends its iTable and token to the successor's backup peer. The backup peer will inform other peers that it has replaced the former immune peer.

5. Simulation Results and Analysis

Table 2. Parameters in Simulations

Parameters	Definitions	Values
N	The scale of network	1000
F	The number of files	10000
γ	The exponent of power-law	2.0
a	Weighted factor	0.5
β	Punishment factor	1.5
P	The ratio of immune peers	10%
K	The ratio of immune resource peers	40%
T	Simulation cycles	30

We use Peersim, a classic P2P network simulator, to simulate the Gnutella network. In the simulations, the parameters are set as shown in table 2. Assume that there are two types of peers: normal peers and malicious peers. The ratio of normal peers is 70% while the ratio of malicious peers is 30%. The ratio of polluted files is 50% which means the network is under a severe polluted environment. Every peer only can download one file at one time. After downloading, normal peers will send feedback and delete the file if the file is detected to be malicious while malicious peers will send the malicious feedback and share the malicious file. The following simulations will testify and analyze the performance of our strategy, the availability under different ratios of malicious peers and recognition rate of malicious files.

5.1 Compare of Whole Performance

The first simulation set is as the initial set, we compare TrustIs with DeepCure immune strategy to analyze the availability of our strategy. In the simulation results, we use NFUR (Normal File Upload Rate) as the KPI and $NFUR = \frac{\text{downloading times of normal files}}{\text{downloading times of all the files}}$. As shown in Figure 2, before 12th cycle, DeepCure's performance is better than TrustIs since DeepCure has enough immune resources at the start phase. But as the simulation goes on, the immune peers has converged more immune resources from resource peers which improves the degree to filter out polluted files. So generally, TrustIs is better than DeepCure on the whole performance.

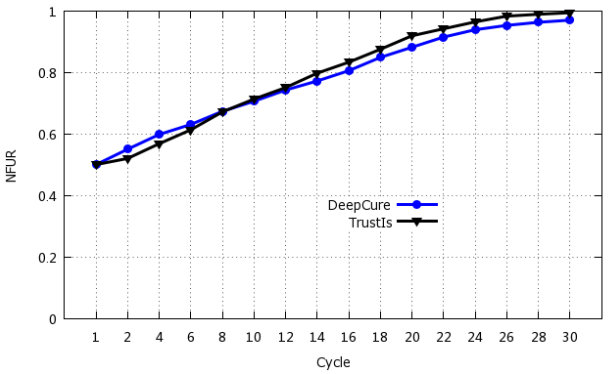

Figure 2. The Evolution of Downloading Rate of Normal Files

5.2 Availability of Strategy Under Different Ratios of Malicious Peers

From Figure 3, we can see that as the number of malicious peers increases, DeepCure gets an obvious drop on the downloading rate of normal files. As the ratio of malicious peers reaches 70%, the NFUR of DeepCure has dropped to 0.047. That is because the weakness of DeepCure on resisting malicious peers. Too many malicious peers will absolutely come across the situation that some malicious peers become the immune peers and thus spread the polluted files. However, in TrustIs, we select immune peers not only with high degree and betweenness but also with high reputation which is more important. Peer reputation can effectively isolate the malicious peers out of the immune sub-network which will improve the efficiency of filtering polluted files.

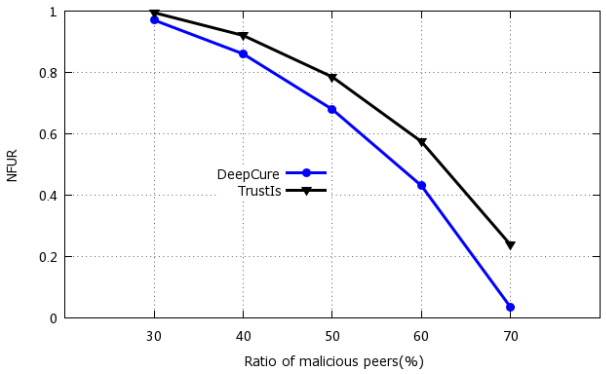

Figure 3. Evolution of Downloading Rate of Normal Peers Under Different Ratios of Malicious Peers

5.3 Recognition Rate of Malicious Files

In this simulation, we use PFFR (Polluted File Filtered Rate) to represent the filtered rate of malicious files while $PFFR = \frac{\text{number of detected polluted files}}{\text{number of polluted files}}$. As shown in Figure 4, TrustIs's PFFR has almost reached 100% after 18th cycle while DeepCure just reaches 80%. The reason is that TrustIs not only uses peer reputation to select immune peer, but also takes object reputation and index reputation into consideration when evaluating the file or index, this to some extent can resist the propagation of polluted files. Moreover, the regular refreshment of reputation of object and index keeps the immune resources more effective and accurate.

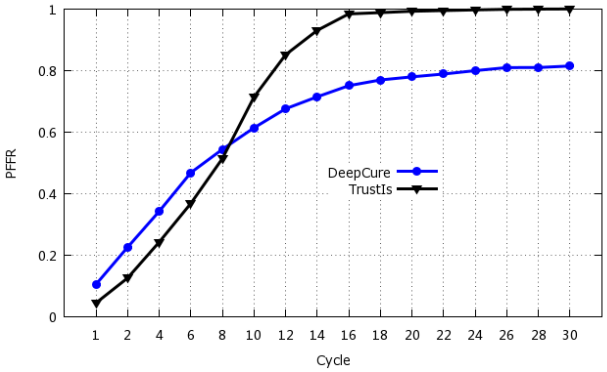

Figure 4. Evolution of Filtered Rate of Malicious Files

6. Conclusion

In this paper, we propose a trust-based immune strategy named TrustIs, which aims to solve the file pollution in P2P network. Previous methods have defects such as large load consumption, low immune efficiency and vulnerability to malicious peers. Our strategy make use of the peer reputation, object reputation and index reputation to guarantee the selection of immune peers and immune resources which improves the efficiency of filtering polluted files and effectively resist the malicious peers. Simulation results show that TrustIs can effectively filter out the files with low reputation and select credible and efficient immune peers to resist the propagation of polluted files.

Acknowledgment

The Fundamental Research Funds for the Central Universities (LGYB201507).

References

[1] J. Mihajlo, F. Annexstein and K. Berman. Modeling peer-to-peer network topologies through small-world models and power laws. IX Telecommunications Forum, TELFOR, **(2001).**
[2] J. Liang, R. Kumar, Y. Xi, K. W. Ross. Pollution in P2P file sharing systems. INFOCOM 2005. 24th Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings IEEE. vol. 2. IEEE, **(2005)**.
[3] W. Dong, S. B. Yang, and X.Q. Liu. Artificial immunology based anti-pollution P2P file sharing system. Grid and Cooperative Computing, GCC 2007. Sixth International Conference on. IEEE, **(2007).**
[4] Z. Y. Huang, X. L. Shi, X. C. Zhou, X. L. Chen. Polluted File Detection Method in P2P File-sharing System. Computer Engineering, vol. 38, no. 13, **(2012),** pp. 102-104.
[5] C. Costa, V. Soares, J. Almeida, V. Almeida. Fighting pollution dissemination in peer-to-peer networks. Proceedings of the 2007 ACM symposium on applied computing. ACM, **(2007).**
[6] W. Kevin, E. G. Sirer. Experience with an object reputation system for peer-to-peer filesharing. NSDI, **(2006).**
[7] C. Cristiano, J. Almeida. Reputation systems for fighting pollution in peer-to-peer file sharing systems. Peer-to-Peer Computing, 2007. P2P 2007. Seventh IEEE International Conference on. IEEE, **(2007).**
[8] C. Reuven, S. Havlin, and D. B. Avraham. Efficient immunization strategies for computer networks and populations. Physical review letters 91.24 **(2003)**: 247901.
[9] W. P. Guo, X. Li, and X. F. Wang. Epidemics and immunization on Euclidean distance preferred small-world networks. Physica A: Statistical Mechanics and its Applications 380 **(2007)**, pp. 684-690.
[10] M. Nilly, T. Kalisky, R. Cohen, D. B. Avraham. Immunization and epidemic dynamics in complex networks. The European Physical Journal B-Condensed Matter and Complex Systems 38.2 **(2004)**, pp. 269-276.
[11] X. F. Meng, and W. Q. Cui. Research on the immune strategy for the polluted file propagation in structured P2P networks. Computers & Electrical Engineering 38.2 **(2012)**, pp. 194-205.
[12] D. L. Duan, R. J. Zhan. Evolution mechanism of node importance based on the information about cascading failures in complex networks. **(2013)**, pp. 68902-068902.
[13] L. X. Huang, F. T. Zou, and F. Y. Ma. Targeted local immunization in scale-free peer-to-peer networks. Journal of Computer Science and Technology 22.3 **(2007)**, pp. 457-468.

[14] Leibnitz, K, Hoßfeld, T, Wakamiya, N, and Murata, M. On pollution in eDonkey-like peer-to-peer file-sharing networks. Measuring, Modelling and Evaluation of Computer and Communication Systems (MMB), 2006 13th GI/ITG Conference. VDE, **(2006).**
[15] S. Huang, H. F. Cui, and Y. M. Ding. Evaluation of node importance in complex networks. arXiv preprint arXiv:1402.5743 **(2014)**.
[16] X. F. Wang, X. Li, and G. R. Chen. The theory and application of complex network. Tsinghua University, Beijing, China **(2006),** Japan.

