A P2P Reputation Model Based on P2P File-Sharing Behavioral Characteristics

Chao Xin, Hui-jun Han, Xiu-qin Lin, and Geng-yu Wei

Lab of Computer Network and Information Security, BUPT, Beijing, China
Key Laboratory of Trustworthy Distributed Computing and Service (BUPT),
Ministry of Education, Beijing, China
{xinchao,weigengyu}@bupt.edu.cn

Abstract. In recent years, P2P networks have become a focus in the industry. The P2P network is open and anonymous, providing the opportunity for false documents, malicious attacks and other malicious acts, so its trust and security issues have emerged. Building reputation model in the P2P network environment is an effective way to solve these problems, but traditional reputation model neglect the difference of pollution files. The more popular of pollution file, the more damage it will bring to P2P network. Tosolve this problem, a P2P reputation model based on P2P file-sharing behavioral characteristics is proposed. By analyzing the P2P file-sharing behavioral characteristics, determine the file's propagation degree. Usethe file's propagation degreeto calculate the files'evaluation value and nodes' reputation value.By this means, increase the reliability of the reputation model, so as toimprove the P2P network security.

Keywords: P2P network, reputation model, file-sharing behavioral characteristics, network security.

1 Introduction

Due to P2P network's anonymity and open characteristics, it's difficult to find an effective and feasiblemethod tosolvethe P2P network securityissues. Traditional network security technologies are also difficult to apply to the P2Pnetwork. How to establish a reasonable and effective reputation mechanism in P2P networks become a hotresearch scholars [1].

Traditional P2P reputationmodeltendto onlyconsider the quality of the documents (authenticity / contaminated) and quality of service, while neglecting the influence of pollution files. This paper proposed a P2P reputation model based on the user file-sharing behaviorcharacteristics. According to analyze the downloading features ofP2P network users, we combine the evaluation value of single file and the propagation degree to determine the node's reputation value. If sharing a pollution file and this kind of files is very popular in P2P network then this node's reputation value will decrease sharply.

2 Analysis of Users' File-Sharing Behavior Characteristics

In order to analyze users' file-sharing behavioral characteristics,we have collected some download information from some main download tools' operator in some large city .The download information includes users' download time, download files'

Y. Yuan, X. Wu, and Y. Lu (Eds.): ISCTCS 2012, CCIS 320, pp. 373–379, 2013.

information, the users' IP information. Based on these users' download information, it is helpful to analyze the P2P network users'file-sharingbehavior characteristics [2, 3] and thereby define a file's propagation influence degree.

2.1 Time Characteristics of Users' File-Sharing Behavior

First, we analyze the relationshipbetweenthe download requests number of P2P network usersand the download time. Results as shown in Figure 1:

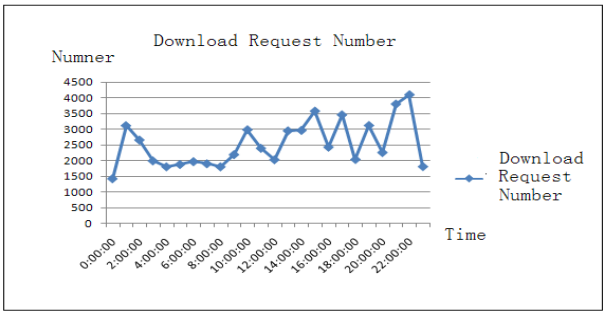

Fig. 1. Download Request Number inDifferent Time

According to the results, we can clearly see that the users' downloads number is in the day of the cycle fluctuation model. Download request number reach a peak at 21: 00 to 23: 00, accounted for 13% of the daily download share.Download request numbermeet the minimum at 00:00.From 03: 00 to 08: 00 users' download request number is relatively stable, fluctuating gently.

2.2 Type Characteristic of Users' File-Sharing Behavior

The file types including executable files, audio files, video files, document files, graphic file and compressed files. According to the statistics of user download data, the file type distribution is shown in figure 2.

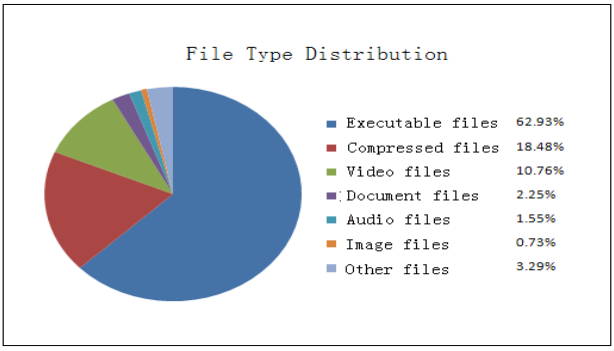

Fig. 2. File Type Distribution

Executable files including EXE file. Compressed file including RAR and ZIP files. Video fileincluding RMVB, AVI, WMV, and RMfiles. Document file including PDF, DOC and TXT files.Audio file including MP3 and WMA files. Image files including JPG and PNG files. It can be seen from Figure 2, the executable file downloads accounted for 62.93% of the total download number.

3 Reputation Model Based on the Users' File-Sharing Behavior Characteristics

This paper presents a P2P network reputation model based on users'file-sharingbehavioral characteristics, including file evaluation value and node reputation value measurement and maintenance. File evaluation value depends on thefile's propagationinfluence degree (i.e.higherfile's propagation influence degree means user will download this file more times).

3.1 Measure of File Evaluation Value and Node Reputation Value

In the traditional model of reputation, the node's reputation value was based on the history score of their servicesquality. In order to gain higher credit, node may share multiple files.A high reputation node can use only onepollution file to undermine P2P network security.In order to reduce this inconsistency behavior, a file's evaluation values will be independent of each files' evaluation values.A file's evaluation value will notbe affected by other files. The node's reputation value is determined by all the files' evaluation values in this node.

The File's Propagationinfluence Degree

Each file has its own files spread influence degree. Files'propagation influence degree can reflect the files' transmission level. After collectinga period of time of users'downloaddata,wecansummarize the characteristics of the user file-sharing behavior for this period of time, then obtained various types of file download ratio and the user downloadsratio in each period of time. Combining these two factors cancalculate the file's propagation influence degree.

Assume that a period of time P2P network users to share files collection isfile={f1 , f2 ,…,fm}, Type (k) representation of the file K's type,| Type (k) | means the download number of this type. Time (k) represents the file K's download request timestamp, | Time (k)| means the download number in this timestamp hour. Then through the statistics of the file type download radioand the filerequest time download radio to determine the file's propagation influence degree β.

The formula is as follows:

$$\beta = \frac{|\mathrm{Type(k)}|}{|\mathrm{File}|} * \frac{|\mathrm{Time(k)}|}{|\mathrm{File}|} * 100 \quad (1)$$

The Measurement of Node's Reputation Value

In each interaction request,request node will score eachserver node. Evaluation score is defined in [-1,1]. It contains the request node's judgments of file'sauthenticity, and satisfaction with the quality of service. A higher value indicates the file's authenticity higher and the higher satisfaction of service. When a node i in a P2P network share a new file j, the initial evaluation valueof this documentis 0:$T_i^0(j)=0$.File evaluation value is updated periodically, if in the nth cycle, the nodes, which downloaded this file, are C^1, C^2,…,C^n, and$S_i^n=\{C^k,\ 1\le k\le n\}$,then evaluation value of the file:

$$T_i^n(j)\begin{cases}\alpha T_i^{n-1}(j)+(1-\alpha)\dfrac{\sum_{k\in S_i^n}Score_k(j)}{|S_i^n|}\beta, & n>1\\ \dfrac{\sum_{k\in S_i^n}Score_k(j)}{|S_i^n|}, & n=1\end{cases} \tag{2}$$

In the formula,$T_i^{n-1}(j)$ is the evaluation value in the N-1 cycle,α is attenuation factor, and $\alpha\in(0,\ 1)$. The smaller theαvalue means this file draw more attention. Use attenuation factorcan prevent the node received in advance to use the real file and thenspread the pollution file.β is the f file's propagation influence degree. In the end of the nth cycle, the evaluation value of file j in node iis$T_i^n(j)$.

The Measurement of Node Reputation Value

Any two files' evaluation values are independent of each other. Node credit (Rep) must reflect this node's contribution degree to the whole network. This includes: shared file number, service number, the authenticity of the files,quality of services, and whether the node's service is consistent. In this paper, the initial reputation of the node is defined as $Rep_{init}=0$, Assuming that node i shared file collection for $File_i=\{f^1,f^2,\dots,f^m\}$,$DL_i(k)$ is download number of the file K. Node i current reputation value calculation formula is as follows:

$$Rep_i=\begin{cases}\sum_{k\in File_i}(T_i(k)*DL_i(k)), & |File_i|>1\\ T_i*DL_i, & otherwise\end{cases} \tag{3}$$

3.2 Maintenance and Update of the File Evaluation Value and Node's Reputation Value

The Query of Document Evaluation Value and Node Reputation Value

When a node i requests a file K, assuming that node j1，j2,…,jm provides this file's download service. This node will broadcast the file K's evaluation value query request <"request"|IDi|k|ID1|ID2|…|IDm|>, and nodej1，j2 ,….,jm will respond to this message <"Reply"|IDj|k|Tj(k)|>after received the request, and then node i will sort the service node according to the service node evaluation value, choose the higher file's evaluation value of n nodes to download.

The Update of Document Evaluation Value and Node Reputation Value Update Assume that node I download a file k from node j, the requesting node i will broadcast filescore information <"score"‖IDi‖IDj‖socorej(k)>, aftermaintenance node of node j receives this information it will update the corresponding document evaluating value and node reputation value.

4 Simulation Results

We use two kind of P2P reputation model and compare the pollution files' propagation to verify the results. One reputation model is based on the P2P file-sharingbehavioral characteristics and the other one is based on trust and recommendation [4, 5]. To verify the Simulation results, we using Java to build an unstructuredP2P simulation platform.

The initial condition of Simulation: there are 1000 nodes in P2P network, each node store 1 to 5 files. In these files, some files are randomly marked to represent pollution files, so as to observe these pollution files' propagation.

4.1 The Compare of Pollution Files' Propagation Degree

We observed the files' propagation in 100 round of simulation. We assume that the number of nodes which have pollution files is 10.each round we let 1 to 9 nodes to try downloading that pollution file.

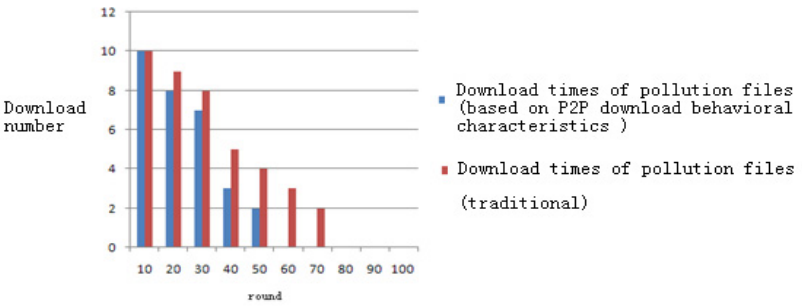

Fig. 3. Compare of pollution files' propagation degree

Figure 3 shows that the download number of pollution files in two kinds of reputation model. We can clearly see that the download number of pollution files in our reputation model (blue bar) is significantly less than traditional reputation model (red bar). That means the pollution files' propagation has been inhibited.

4.2 The Compare of Node's Successful Download Radio

A successful download means this node download a healthy file. We compare the successful download radio in each reputation model to verify the reliability of the reputation model.

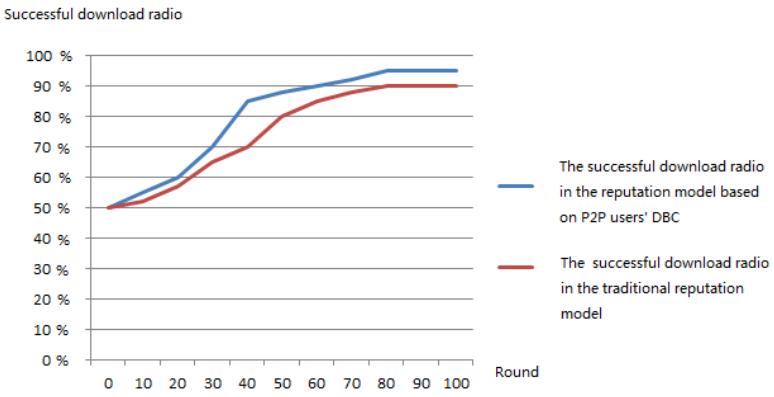

Fig. 4. Compare of node's successful download radio

Figure 4 shows that the successful download radio in our reputation model is higher than traditional reputation model. That means the successful download radio has been improved. Users are more likely to download healthy files.

4.3 Result analysis

Form figure 3 and figure 4, we can see that the pollution files' propagation has been inhibited and the successful download radio has been improved. Because the reputation model based on P2P file-sharingbehavioral characteristics use the files'propagation degree to determine the files'evaluation value. That makes the node's reputation value sharply decrease when this node share a popular pollution file. In that way, we can improve the reliability of the reputation model.

5 Conclusion

The reputation model based on the P2P file-sharingbehavioral characteristicscombines the files'evaluation value in server node and the files'propagation influence degree to choose whether download from this sever node. Thehigherpollution files'propagation influence degree, the more nodes' reputation value and files' evaluation value will be decreased. By this way, thispollution file's propagation will be inhibited and the P2P network security will be improved.

References

1. Kui, L., Dong, L.S.: Studies on Reputation Model in P2P Networks, pp. 1–2 (2008)
2. Chen, B.-G., Xu, Y., Hu, J.-L., Zhang, L.: Researchon User Behavior Characteristics of P2P File Sharing Systems. Computer Science 34(12), 122–142 (2007)
3. Huang, Z.-H., Lu, S.-N.: Peer to Peer Model Based on User Behavior. Computer Engineering 37(11) (2011)
4. Xi, J., Wang, Y., Lu, J.-D.: P2P Reputation Model Basedon Trust and Recommendation. Computer Engineering 35(4), 143–145 (2009)
5. Cornelli, F., Damiani, E., Capitani, S.D.: Choosing Reputable Serversin a P2P Network. In: Proc. of the 11th International World Wide Web Conference, Honolulu, Hawaii, USA: [s.n] (2002)