The Sixth International Symposium on Operations Research and Its Applications (ISORA'06)
Xinjiang, China, August 8–12, 2006
, pp. 319–326

P2P Contents Distribution System with Routing and Trust Management

Hiroyuki Kawano*

Department of Information and Telecommunication Engineering, Nanzan University, Japan

Abstract The "peer-to-peer" architectures are characterized by the direct sharing of computer resources without centralized servers, and many researchers try to improve the performance of P2P network systems. In this paper, firstly, in order to stabilize network connections, we describe our proposed query routing mechanism based on collaborative information filtering. Secondly, we discuss the architecture of hierarchical query routing scheme, and thirdly we consider the reliability of P2P contents distribution systems by using trust management. Since it is important to improve the stability and robustness of P2P systems. Our trust management algorithm is based on public-key cryptosystems and Bayesian probabilistic functions.

1 Introduction

At present, the volume of information all over the world is rapidly increasing, it is possible to store and analyze PB (10^{15}) and EB (10^{18}) scaled data in some practical applications. Actually, in order to share and transfer huge volume of information resources, various types of distributed systems over peer-to-peer (P2P) network have been developed. P2P applications are extremely important in the view points of scalability and flexibility. Therefore, many researchers and developers try to improve the functionalities of routing mechanisms, search systems, trust managements and many others by using various technologies [1].

Firstly we focus on the P2P system with query routing protocol in this paper. We describe our proposed P2P contents distribution system with routing tables based on information filtering techniques. Our proposed architecture was implemented by JXTA, which was developed by Sun Microsystems [12]. JXTA provides flexible XML routing descriptions called as Query Routing Protocol (QRP), but it has some weakness of handling metadata in query messages and query processing.

In our previous works [8, 9], we try to resolve several problems and propose the topic-driven routing system which utilizes standardized metadata (e.g MPEG-7, Dublin Core) of information resources. Our proposed QR (Query Routing) protocol is based on the topic-oriented query processing. QR protocol reduces the number of

*E-mail: kawano@it.nanzan-u.ac.jp

flooding messages in P2P network. However, QR stores the attributes of information resources at full length, the size of query routing tables is becoming large easily.

Secondly, by using generalization techniques of data mining to QR protocol, we propose HQR (Hierarchical Query Routing) [7], which has the hierarchical structures of routing tables. We also evaluate the performance of QR and HQR depending on the different levels of hierarchical routings.

Thirdly, we focus on the important problems of trustness. Even if one malicious peer client exists in the network, the performance of collaborative filtering algorithms like QR or HQR decreases gradually. It is very hard to keep the quality of information routing tables clean. Therefore, in P2P network, reputation, trust, security, reliability and related concepts are also important for effective sharing files and realizing robust contents distribution.

Thus, we discuss the trust problems, and we propose trust management methods based on public-key cryptosystems [5] and Bayesian probabilistic functions, in order to implement trust mechanisms in our contents distribution system. We further evaluate the performance of trust and reputation chains, which play important roles to guarantee high reliability of our system.

In this paper, we discuss the characteristics of our routing management in Section 2 and trust chain scheme in Section 3, from several view points including information filtering, recommendation, trustness, chain of trust networks and others.

2 Routing Management in P2P Content Distribution Systems [7, 8, 9]

In this section, we present our proposed routing algorithm based on the classifier and the distiller, we also apply hierarchical management to our algorithm in order to realize a topic-driven search mechanism in P2P content distribution system in Fig.1.

2.1 P2P Routing using Classifier and Distiller

The classifier learns topics of interest from the user's stored information resources, and focuses on a set of interesting peer hosts, which have similar preferences. On the other hand, the distiller is executed to discover hub hosts which have higher similarity of interesting peer hosts. This learning process is similar to the process of web structure mining, such as PageRank and related algorithms. The major steps are described in the followings.

1) Building index files and scoring hosts

Firstly, we produce an advanced index file which describe the characteristics of files stored in a host in order to realize the topic-driven search. Actually, the index file contains a lot of attributes, such as name, extension, size, date, creation system of the files and so on. In addition to those primary attributes, we may also utilize the anchor tags in hypertext files and the number of referenced files from web pages.

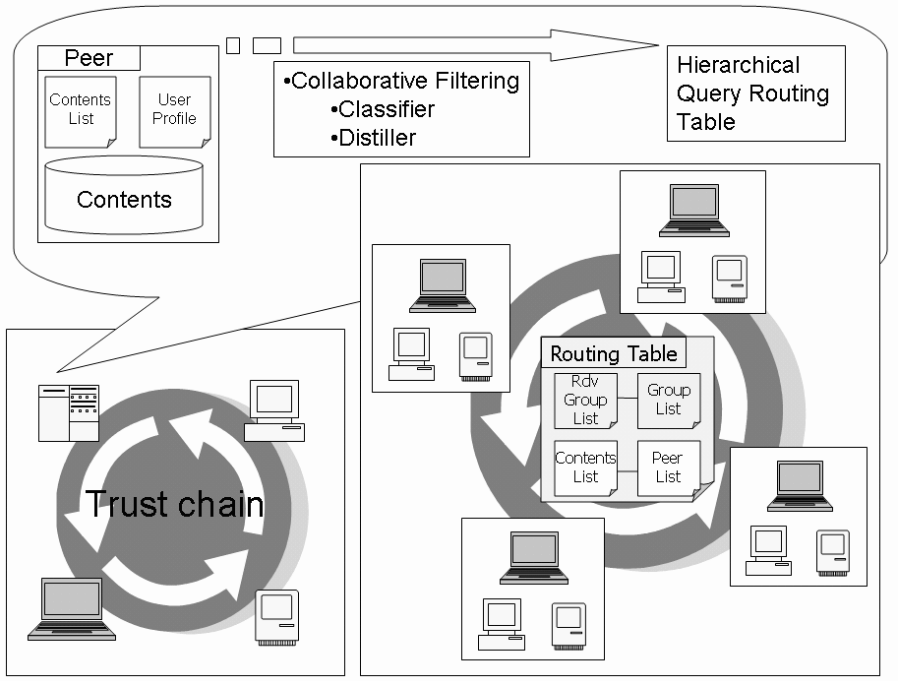

Figure 1: Concept of P2P contents distribution with routing and trust management

2) Discover a group of valuable hosts using classifiers

Next, we execute the classifier like web crawlings by intelligent robot programs. Our classifier selects hosts considering whether they have optimal resources and whether they are suitable for retrieving resources. The peer host chooses optimal hosts by the classifier and demands the index files of the peer hosts. The classifier is based on the score function of index with metadata, network distance and other attribute values.

3) Distributed index refinement using distillers

The user can select suitable hosts with optimal index files by the classifier. If the user transfer the whole index files or applicable portion from a host, he can distill target files by using these downloaded index files without storing all files in a user's host. Therefore, we introduce hub score as a parameter to estimate how many a host has suitable index files.

Thus, hosts with large hub scores have possibilities of handling suitable index files. We can choose optimal hub hosts by hub score. The query host can execute search more efficiently by using a hub host which contains many more suitable index files.

2.2 Hierarchical Query Routing

P2P file-sharing systems are classified into following four types.

i) Client-Server type controlled by super-peers
ii) Pure P2P system using flooding messages
iii) Structured P2P system with routing index scheme
iv) P2P system with query routing messages

In this subsection, we apply QR algorithm and HQR algorithm to the fourth P2P system [7].

Query description based on resource characteristics

The topic driven query routing requires metadata of information resources to analyze resource characteristics. In our previous studies, we applied standardized metadata (e.g., MPEG-7) to QRP (Query Routing Protocol) of JXTA search. Typical metadata attribute of MPEG-7 are as follows:

1. MPEG-7: resource information
2. CreationMetaInformation: information of data creation
3. MultimediaContent: format, extension and so on
4. UsageMetaInformation: author's copyright and other properties

QR adopts a query routing system in order to reduce the number of broadcasting query messages. However, QR has a problem that the amount of storage for routing table will be increasing monotonously.

Therefore, we propose generalizing routing table based-on the hierarchical query routing mechanism (HQR). QR only stores the attributes of data and the peer-addresses at full length. On the other, HQR aggregates the attributes of data into the hierarchical structure and stores peer-addresses associated to only neighbors.

Concept of HQR

1. The consumer node C makes requests, and C checks the query in his own registers, and it forwards the query to suitable peer A.
2. The peer A receives the query and checks the query-uid (query-unique ID). If A has never received the query-uid, A checks the query in his own registers including hierarchical routing information. Moreover if he finds out a suitable resource in his registers or the hierarchical routing table, A sends back response message to C and forwards a query to the suitable peer P. If the query does not match the information in his registers, A forwards a query to its neighbor and reduces TTL of the query. This operations are repeated until $TTL = 0$.
3. When C receives a response message, it checks the resource information and stores it in his query routing table. This resource information provides routing operations effectively.

The tradeoff between the cost of query forwarding and the amount of storage for routing information have been studied in [7].

3 Trust Management in P2P Content Distribution System [5]

P2P anonymous file sharing systems without centralized servers cause various problems from the view point of security, reliability, reputation management, trustness, risk and others [3].

In this section, we present that our proposed algorithm constructs appropriate trust networks of peer clients by using the relationships of trust values with public-key cryptosystems. Furthermore, the multiple trust chains with Bayesian probabilistic functions also provide good characteristics of reputation in our system.

3.1 Trust Management of P2P systems

Actually, in popular web systems, the framework of PKI is sometimes utilized丐For example, in the online shopping sites, social networks based on transaction data play an important role of recommendation. In order to improve reliability of web systems, many related technologies and frameworks including PGP (Pretty Good Privacy) have been proposed and implemented. Essentially, those reliability strongly depend on various aspects of the system architectures, system availability, access control of files and resources, reliability and security of contents, authentication of data providers [3]. In addition to above technical trustness, social trustness are also important factors.

Generally, users have different preference vectors including constraints of users' interestingness, quality of contents, type of media and waiting time of transfer, higher reliability is required by users in fully distributed P2P systems. Therefore, in Table 1, we show various trust management frameworks have been proposed by many researchers [2, 3].

Table 1: Technical aspects of P2P trustness

Methods		Techniques
Profile		Reliable attribute values provided by peer host [11]
Recommendation		Recommendation values provided by other hosts [2]
Reputation	Average of values	Average values of behaviors
	Probabilistic model	Utilization of Baysian approaches [13]
	Trust ring	Utilization of trust chains [4]
	Flow model	Flow model of reputation in P2P entire system [6]

There are various ways to improve trustness in P2P systems, for example, trustness is managed by using DHT in [6], different evaluation functions are proposed in [14], Bootstrap servers are utilized in [10]. The propagation of web trustness is discussed in the paper [4], we also try to introduce the mechanism of trust chain in our proposed P2P content distribution system.

3.2 Concepts of Trust Management

In our works, we would like to implement our contents distribution system with low load traffic and high trust, by using public-key cryptosystems and a signature scheme in P2P networks. In our paper [5], we described detail algorithms and trust tables of content request/provide peers.

Trust value of peer B himself by peer A is presented by $C_P(A,B)$, Trust value of peer B's public key by peer A is defined as $C_K(A,B)$. Then, we define the trust value of peer B by peer A as $pre\_C_P(A,B)$, using system parameter α in trust chains,in (1).

$$C_P(i,j) = \max(\alpha C_P(i,k)C_P(k,j), pre\_C_P(i,j)) \tag{1}$$

If there exist multiple trust chains in the network, $C_K(A,B)$ is maximum values among all trust propagating networks.

The following equations give updation of trust values. (a,b,c,d) are parameters for trust updating values, (K,L,M,N) are normalized parameters with $[0,1]$, and the repetition of transfer Success/Failure is x.

Success:

$$C_K() \leftarrow C_K() + Ka^x \tag{2}$$
$$C_P() \leftarrow C_P() + Lb \tag{3}$$

Failure:

$$C_K() \leftarrow C_K() - Mc^x \tag{4}$$
$$C_P() \leftarrow C_P() - Nd^x \tag{5}$$

3.3 Characteristics of Trust Propagation

We have simulation studies of propagation of trustness in order to validate the effectiveness of our trust management scheme. Firstly, we evaluate the size of trust tables with different request frequency. Actually, peer hosts having similar files frequently communicate each others. Those hosts also have similar attribute values in their routing tables. Based on our simulation results, the size of trust tables is sufficient within 5% of number of connecting peers, and it is possible to keep trust chain with more than 30% of peers joining to the network.

We also discuss the characteristics of our reputation management and trust chain scheme, from various view points including recommendation, simple evaluation of trust values, property of trust networks and other approaches. We have to improve the performance evaluation models for reputation mechanism more theoretically. Various researches, which are related to trust and reputation, such as resource trading models, economical models, game theoretic analysis and others, are useful and helpful to discuss the advanced problems.

4 Conclusion

In this paper, in order to stabilize network connections, we proposed query routing mechanism based on collaborative information filtering, and we apply hierarchical generalization to routing algorithms firstly.

However, even if one malicious peer client exists in the network, the performance of collaborative filtering algorithm decreases, moreover it is hard to keep the trustness of information routing tables. Therefore, in our following work, we discuss the problems of trust management, and we propose trust management based on public-key cryptosystems and Bayesian probabilistic functions, in order to implement trust frameworks in our contents distribution system.

In the future, we need to improve the performance evaluation models of trust management more theoretically.

Acknowledgments

A part of this work is supported by 2006 Nanzan University Pache Research Subsidy I-A-2.

References

[1] S. Androutsellis-Theotokis and D. Spinellis. A survey of peer-to-peer content distribution technologies. *ACM Computing Surveys*, 36(4), 335–371, 2004.

[2] K. Chopra and W. A. Wallace. Trust in electronic environments. *Proc. of the 36th Hawaii Int. Conf. on System Sciences*, 2003.

[3] N. Daswani, H. Garcia-Molina and B. Yang. Open problems in data-sharing peer-to-peer systems. *Proc. of the 9th Int. Conf. on Database Theory*, 1–15, 2003.

[4] R. Guha, R. Kumar, P. Raghavan and A. Tomkins. Propagation of trust and distrust. *Proc. of the 13th Int. World Wide Web Conf.*, 403–412, 2004.

[5] Y. Ito and H. Kawano. Reliable trust function with cryptographic chains in peer-contents distribution system. *Proc. of DBWeb'05*, 133–140, 2005. (in Japanese)

[6] S. Kamvar, M. Scholsser and H. Garcia-Molina. The eigentrust algorithm for reputation management in P2P networks. *Proc. of the 12th Int. World Wide Web Conf.*, 2003.

[7] H. Kawano and H. Kishi. Hierarchical query routing in P2P information filtering systems. *Proc. of the Fifth International Symposium, ISORA'05*, 217–226, 2005.

[8] M. Nakatsuji, M. Kawahara and H. Kawano. Advanced index refinement by classifiers and distillers in P2P resource discovery. *International Conference on Intelligent Agents Web Technology and Internet Commerce' 2003*, 272–285, Vienna, Austria, Feb. 2003.

[9] M. Nakatsuji, M. Kawahara and H. Kawano. The architecture and performance of topic-driven P2P resource discovery system. *IEICE Transaction*, J87-D1(2), 126–136, Feb. 2003.

[10] A. Singh and L. Liu. TrustMe: anonymous management of trust relationship in decentralized P2P system. *Proc. of the 3rd Int. IEEE Conf. on Peer-to-Peer Computing*, 2003.

[11] N. Stakhanova, S. Basu, J. Wong and O. Stakhanov. Trust framework for P2P networks using peer-profile based anomaly technique. *Proc of the 2nd Int. Workshop on Security in Distributed Computing Systems*, 203–209, 2005.

[12] S. Waterhousek. *JXTA Search: Distributed Search for Distributed Networks*, Sun Microsystems, Inc., California, 2001.

[13] A.Withby, A.Josang and J.Indulska. Filtering out unfair ratings in Bayesian reputation systems. *Proc. of the 7th Int. Workshop on Trust in Agent Societies*, 2004.

[14] L. Xiong and L. Liu. PeerTrust: supporting reputation-based trust for peer-to-peer electronic communities. *IEEE Transactions on Knowledge and Data Engineering*, 16(7), 843–857, 2004.