Cluster-Based Reputation Model in Peer-to-Peer Network

Mei Chen, Kenji Kita, and Xin Luo

***Abstract*—Nowadays, Peer-to-Peer network represents a large portion of internet traffic, and becomes fundamental data sources. Because of lacking the security mechanism from third-party, P2P network will face some severe trust problems such as service faking and resource abusing by some malicious peers. The conventional security measures can not be used to cater for this demand, whereas the scenario based on reputation has widely been accepted. Through studying the present reputation, the paper presents a cluster-based reputation model (CBRM). The model is consisted by reputation mechanism and cluster. In the model, we take the reputation mechanism for realizing the security transaction; and the network topology structure of CBRM adopts the cluster, so efficiency of reputation management is noticeably raised. In order to improve security, reduce the network traffic brought by management of reputation, and enhance stability of cluster, when we select reputation, the average historical online time, and the network bandwidth as the elementary components of the comprehensive performance of node. Simulation results showed that the proposed model improved the security, reduced the network traffic, and enhanced stability of cluster.**

***Index Terms*--P2P; reputation; trust; security; cluster; network;**

I. Introduction

Peer-to-Peer (P2P) networks are self-configuring networks with minimal or no central control [1]. It integrates the scattered network resources, improves the capability of resources sharing, and maximizes the utilization of resources. So P2P network gets the fast development. Due to the open, free, and anonymous nature of P2P network, so it is more vulnerable to dissemination of malicious or spurious content, viruses, malicious code, worms, etc. The traditional security techniques developed for their centralized distributed systems are inappropriate for P2P networks by the virtue of the centralized nature. For these problems, one feasible way to minimize the threats is to establish the reputation model.

In recent years, reputation has widely been studied in P2P network. Both reputation and trust are different, but also have relations both. Trust is abstract and has the intuitive sense [2], the subjectivity, dynamic; and reputation is metric of trust and easily builds model.

Reputation model is based on a network topology and builds the reputation mechanism. Resnick et al. [3] defines the reputation system as “a system that collects, distributes, and aggregates feedback about consumer's past behaviors.” Reference [4], for the problems of self-replicating inauthentic files, presents a distributed and secure method to compute global trust values, based on power iteration. Literature [5], in order to use community-based reputations to help estimate the trustworthiness of peers to minimize threats of communities, presents a reputation-based trust supporting framework. Literature [6], for enabling peers to represent and update their trust in other peers for sharing files, proposes a Bayesian network-based trust model for building reputation based on recommendations in P2P networks. Literature [7], for authentication and recommendation of trust, presents a method for the valuation of trustworthiness which can be used to accept or reject an entity as being suitable for sensitive tasks. Literature [8] poses the reputation model based on reinforcement machine learning from a computer-science perspective. Literature [9], for the problems caused by anonymity in P2P, proposes a self-regulating system where reputation sharing is realized through a distributed polling algorithm.

But the present models emphasize particularly on realization of function and take little account of feasibility of application, especially in the aspects of network security, stability, and network traffic. Combined with the existing reputation model and features of the P2P network, the paper proposes the cluster-based reputation model. The model is consisted of reputation mechanism and cluster. Reputation mechanism realizes the network security by reputation evaluation. Cluster is a new network topology, its organization structure is flexible, and its management is very convenient. In the model, reputation represents node’s credibility. For a node, the higher the reputation of node is, the more credible it is. In the election of super node and transaction between nodes, node of the high reputation will have more priorities or chances, and the low reputation will have more restrictions. Because the comprehensive performance of node considers the average historical online time and network bandwidth, the network flows and stability of cluster are improved.

II. Cluster of CBRM

A. Concepts of cluster

In recent years, the hybrid P2P structure [10] is becoming a large portion of P2P network. In conventional hybrid P2P structure, super nodes may overload such as query routing and file management, reputation management, the difficulty of the management is great, and network traffic is greatly increasing. This is because node performance is influenced by many factors such as reputation, the average historical online time, and network bandwidth, etc, so the oversimplified way of selecting a super node may not match

Manuscript received September 13, 2011, revised September 16, 2011. This research was partially supported by “Grant-in-Aid for Scientific Research (B) (21300036) from Japan Society for the Promotion of Science”.

Mei Chen and Kenji Kita are with the University of Tokushima, Tokushima, Japan. (e-mail: chen-mei@iss.tokushima-u.ac.jp; kita@is.tokushima-u.ac.jp).

Xin Luo is with the School of Computer Science and Technology, Donghua University, Shanghai, China (e-mail: xluo@dhu.edu.cn).

between overlay topology and geography locality. For these problems, in CBRM, the paper posed three-tier topology based on cluster. Cluster divides the whole P2P network into a lot of small elementary units. Between clusters are independent. Every cluster is a self-managing basic unit in P2P network, and usually consisted of unique cluster head, unique gateway and a number of common nodes. According to these descriptions, the cluster-based P2P network topology is as Fig. 1.

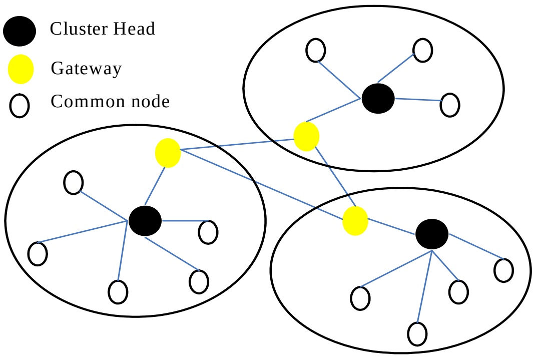

Figure 1. The cluster-based P2P network topology

Common node (CC): it is also called the resource node or terminal node, and can enjoy and share the corresponding level services from P2P network.

Gateway (CG): it is similar to route of network and responsible for communication between clusters, and has the rights of enjoying services and the duty of acting as query routing.

Cluster head (CH): it is the manager of cluster, responsible for managing cluster, and the rights of enjoying services.

Service node: Service node is made up of Cluster head and Gateway.

According to the above definition, the basic structure of cluster is as Fig. 2:

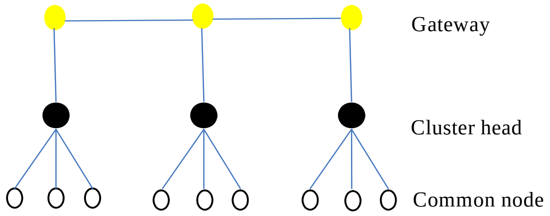

Figure 2. The basic structure of cluster

In a cluster, the comprehensive performance of cluster head is best such as reputation, the average historical online time, and network bandwidth; and next is gateway.

Because of flexible structure of cluster, its application and scalability are very good, management is convenient and efficient, and so network traffic also greatly reduces.

B. Cluster scale

With join of new node, the quantity of members of cluster is increasing. Because the management ability of cluster head is limited, when the quantity of members is over the optimum quantity, and the management efficiency of cluster head will sharply decline.

For the problem, literature [11] posed the scale theorem of cluster. In the theorem, the quantity of nodes of cluster is connected with the performance of cluster. C denotes the set of cluster, $\forall C_i \in C$, scale of cluster C_i is the following: $k_i \le Capacity(C_i) \le 3k_i - 1$, where k_i is determined by overload capacity and bandwidth. When the scale of cluster C_i is over ($3\,k_i$ -1), the cluster will divide; then, when the scale of cluster C_i is below k_i, the cluster will merger. In the theorem, cluster scale is realized by division and merger.

The theorem effectively solved the problem of cluster scale, but division and merger of cluster brought a lot of network traffic. And because of the dynamic of P2P network, the network may greatly have changed after division and merger. So scalability and application of network are poor.

In CBRM, let S_i denote the quantity of average historic nodes of i region; Q denotes the quantity of nodes which cluster head can process and represents the management capability of cluster head. $\min Q_i$ denotes the minimum quantity of the nodes of i region; $\max Q_i$ denotes the maximum quantity; $optQ_i$ denotes the optimum quantity; then, scale of cluster (Q_i) is the following:

$$0 < \min Q_i \le Q_i < \max Q_i \le S_i \qquad (1)$$

$$0 < \min Q_i \le optQ_i < \max Q_i \le S_i \qquad (2)$$

The improved method obviously enhances the stability of P2P network, reduces the network traffic, and improves the scalability of P2P system.

C. Comprehensive performance of node

In P2P network, node performance is influenced by many factors, especially reputation, average historical online time, and network bandwidth. So when we evaluate the comprehensive performance of a node, these factors need to be considered. In CBRM, the comprehensive performance of node is consisted of reputation, the average historical online time, and network bandwidth.

In P2P network, every node is self-organizing and self-managing. Because of the autonomous behaviors, the network is easily vulnerable to kinds of threats from malicious nodes. In CBRM, node behaviors can be categorized into three basic behaviors such as Goodwill, Selfish, and Malicious. Their definitions are the following:

1) Goodwill (G). The transaction service is duly in compliance with demand.
2) Selfish (S). The node rejects offering service or interrupts the ongoing service, etc.
3) Malicious (M). Some nodes join the P2P network to do some destructive activities such as propagating virus, malicious attacks and so on.

For selfish nodes, they just want to enjoy resources, but do not share their resources with other nodes. The phenomenon is free rider or homogenization. Selfish problems have led to many new problems such as network traffic increasing, resources shortage, etc. Then malicious nodes, they mainly destroy the P2P network. For these behaviors, in CBRM, we use reputation for representing node behaviors and reputation mechanism for restricting their behaviors. After transacting, request node will evaluate the service, and resource node will get a reputation evaluation.

In order to evaluate reputation and calculate reputation value, we firstly quantize reputation. Let R denote reputation. The calculation formula of reputation is as follows:

$$G = \sum_{i=0}^{n} g_i \tag{3}$$

where g_i is the goodwill evaluation of the i th of node.

$$S = \sum_{j=0}^{m} s_j \tag{4}$$

where s_j is the selfish evaluation of the j th of node.

$$M = \sum_{k=0}^{l} (-e^k m_k) \tag{5}$$

where m_k is the malicious evaluation of the k th of node. e^k is the punishing factor. According to (3), (4), (5), the formula of R is as follows:

$$R = \begin{cases} 0,(G=S=M=0) \\ \dfrac{G+S+M}{G+S+|M|} \end{cases} \tag{6}$$

R value ranges from -1 to 1. When the node joins the P2P network for the first time, its initial reputation value is equal to 0; when R value ranges from -1 to 0, and it represents that node is incredible; when R ranges from 0 to 1, and it represents that node is credible.

Nodes of P2P network have the feature of dynamic, they frequently join and leave. If the online time of service node is short, the network will greatly fluctuate. So the online time of service node is of vital importance, it represents stability of cluster. In CBRM, we record the past online time of node, and use T to denote the average historical online time of node. The calculation formula of T is as the following:

$$T = \mathrm{E}(t_i) \tag{7}$$

where t_i represents the i th online time of node; T is expectation of t_i.

Network bandwidth is an important influence factor for reputation management. If it is too small, efficiency of reputation management will be low. In CBRM, calculation of the network bandwidth between nodes is from literature [12]. Where W_1 represents bandwidth of cluster inner, and W_2 for cluster outer. N is the total amount of nodes, the optimal quantity of node that cluster head can manage is M, then formula of M is as follows:

$$M = W_1\sqrt{N} / W_2 \tag{8}$$

Then the best connectivity of node is as the following:

$$\delta = N / M = W_2\sqrt{N} / W_1 \tag{9}$$

Let d denote the degree of node, and then its correlation is as the following:

$$B = |d - \delta| \tag{10}$$

B represents performance of the network bandwidth. The smaller B is, the more suitable it serves on cluster head.

According to the above formulas, calculation formula of the comprehensive performance of node is as follows:

$$C = C(R,B,T) = \lambda_R \times R + \lambda_B \times B + \lambda_T \times T \tag{11}$$

$$\lambda_R + \lambda_T + \lambda_B = 1 \tag{12}$$

where R is reputation of node; B is bandwidth of node; T is the average historical online time of node. λ_R, λ_B, and λ_T are their corresponding coefficients. In order to reduce the network traffic and enhance stability of cluster, we let $C0$ represent threshold of C, $R0$ for R, $B0$ for B, $T0$ for T, then $C0$ is equal to $C(R0,B0,T0)$. These thresholds and their corresponding coefficients are got by historical records from P2P network. For candidate of the cluster head and gateway, we demand that their C must be greater than $C0$.

III. Reputation Mechanism of CBRM

Reputation research is mainly aimed at building reputation mechanism. The goal of reputation mechanism would manage reputation of node and predict the future behavior of node by its past behaviors. Reputation mechanism is consisted of reputation information storage, reputation information aggregation, and reputation stimulation.

A. Reputation storage

In order to more accurately predict the future behaviors of node, every time the reputation evaluation of node must store. The commonly used methods store the reputation data in itself or third-party. But these methods will bring a lot of network traffic and the security is low. In CBRM, way of storage is the following.

Every node stores own reputation data evaluated by other nodes, the data will be encrypted and prohibited against modifying. Modification of reputation information is operated by its cluster head.

Through the improved method, operations of reputation storage are done in cluster inner and by cluster head to which the node belongs, so the new way greatly reduces network traffic and enhances the network security.

B. Reputation aggregation

When a node requests a service, it firstly knows about the reputation information of resource node. Cluster head to which resource node belongs aggregates the reputation information of resource node, and sends result to the request node. The calculation formula of reputation aggregation refers to the above reputation.

C. Reputation stimulation

P2P network is a platform of open and free resources sharing. In the network, every node is equal in status and under no restraint from third-party. In order to make more nodes actively cooperate under no restraint from third-party, stimulation is needed. In CBRM, the description of reputation stimulation is as follows:

The higher the reputation is, the more chances the node will get resources. Especially, when cluster is busy or resource is shortage, the higher reputation nodes will have the priorities of obtaining resources.

IV. Realization of CBRM'S Cluster

The cluster has the life cycle, and it includes the initialization of cluster, node join, node leave, disintegration of cluster. In accordance with the life cycle of cluster, the realization of cluster is the following.

A. The initialization of cluster

The main tasks of cluster's initialization are electing the cluster head, and then building cluster. After formation of cluster, cluster head sends invitation of join to the nearby nodes. First of all, the two problems are needed to be considered. One is efficiency of clustering. Because of the dynamic of P2P network, if the efficiency of clustering is too low, and the formation of cluster will cost a lot of time. After the formation, the nearby network may have greatly changed. The other is cost of clustering. In the process of clustering, it will certainly bring a lot of network flows, and that will not be avoided, so we must improve the initialization of cluster and reduce the network traffic.

Based on the above two reasons, initialization of CBRM' cluster takes the dynamic clustering, and the realization is as follows:

In the area of no cluster, for a new node which is ready to join the P2P network, first of all, it sends application of join and waits for response. Because of no cluster in the neighborhood, the new node doesn't get any answer. Then, it will calculate its own comprehensive performance. If the result is greater than $C0$, it is to be cluster head, or just waiting the join invitation of the nearby cluster head. In order to reduce the network flows, the waiting node will be limited to do any activity. After formation of cluster, cluster head will invite the nearby nodes to join.

B. Node join

After the formation of cluster, the nearby nodes will join it. New node firstly sends the join application, after getting the join permission from nearby cluster, and both formally starts to establish communication. After communicating, the cluster head calculates C of the new node and decides its identity. For a new node, its identity may be one of the following three kinds:

1) Cluster head: C of the new node is greater than the current cluster head's.
2) Gateway: C of the new node is greater than the current gateway's, or it is the second node meeting $C0$ in the cluster.
3) Common node: in addition to above two cases, the node can be only the common member.

In order to enhance stability of cluster, for the node join, the paper proposes the concept of virtual backup. Its definition is as follows:

After generating the gateway, cluster head and gateway mutually store the management information of the other party in order to quickly restore when occurring the exception.

C. Node leave

Because of freedom, dynamic, abnormality, node frequently leaves. Node leave may be normal and abnormal. The basic principles which handle the node leave are as follows:

For the normal, node must apply before leaving. But, for the exception, cluster head must communicate with its members at regular internals, if detecting its members which abnormally left, and timely handle. After node leaving, members of cluster may need adjusting. Let C denote the comprehensive performance of node, principles of adjustment are as follows:

If the quantity of nodes meeting $C > C0$ is greater than 2, the greatest one is cluster head and the greater one is gateway; if equal to 1, the node is cluster head and also serves as gateway; less than 1, cluster will disintegrate.

D. The disintegration of cluster

When cluster doesn't have nodes meeting more than $C0$, the left nodes don't meet requirements in the aspects of reputation, the average historical average online, and network bandwidth. In order to reduce the network traffic, we will limit activities of these nodes and let them wait for the join invitation of cluster.

E. Reputation management strategy of cluster

In CBRM, reputation management is realized by cluster head. Main contents are as the following:

1) For choosing the cluster head, C of node must meet $C0$. When transacting between nodes, the requested node firstly considers R of the resource node, and next is C.
2) Cluster head manages the reputation information of its members, including such as query, update, etc. Before transaction, cluster head of service node aggregates its reputation data, and sends result to request node. After transaction, in according to reputation evaluation of request node, cluster head modifies reputation information of service node. For evaluation of malicious behavior, the request node must also offer evidence in order to prevent malicious slander.
3) When node serves as cluster head or gateway, it will be prohibited from providing service of resources sharing, but it can enjoy resource services and have more priorities of getting resource, especially when shortage of resources.

For cluster-based reputation management, all operations are carried out in cluster inner, so management is very convenient, operation is very simple, and network flows brought by management will greatly reduce.

V. Simulation and Experimental Analysis

A. Experiment designs

To test effect of the improved model, the paper performs simulation experiments. Simulation tool uses the PeerSim [13] which is specially designed to simulate the P2P network. PeerSim is the cycle-based engine, to allow for scalability, use some simplifying assumptions, such as ignoring the details, the transport layer, etc.

Through the three indices, namely, security, stability, and network traffic, we perform experiments.

Let S denote the rate of effective download (RED) and it represents the proportion of security transactions in total transactions; N_1 is the times of security transactions; N is the total times of transactions, then calculation formula of S is as the following:

$$S = N_1 / N \tag{13}$$

The greater S is, the more security the system is.

Because cluster head is manager of cluster, so the average historical online time of cluster head represents the stability of cluster. When choosing cluster head in simulation, we also consider the average historical online time, so we can test the stability of the cluster by calculating the online time of cluster head in simulation. Let T denote the average online time of all cluster head (AOT); N is the quantity of cluster head; t_i is the online time of cluster head i. Then calculation formula of T is as the follows:

$$T = \sum_{i=1}^{N} t_i \; / \; N \tag{14}$$

The greater T is and the more stable the system is.

Let F denote the network traffic (NT). NT is determined by data packet which includes query request, response message, file download, connection request, etc. The smaller F is, the less the network flows are.

B. Experiment results and analysis

Figure 3 is the security verification.

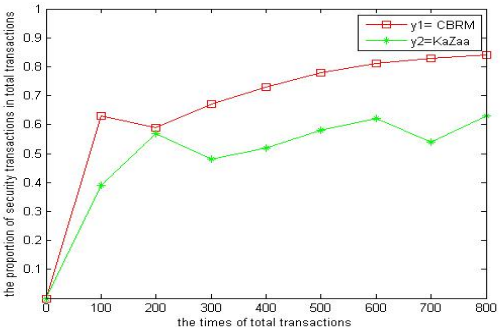

Figure 3. the rate of effective download (%)

In the experiment, the Goodwill nodes account for 70%, the Selfish for 20%, and the Malicious for 10%. At the beginning, RED of CBRM is showing the fluctuation, which is mainly caused by the behaviors of selfish and malicious, but the selfish behaviors are restricted and malicious behaviors are quickly punished, so RED is steadily growing. Because of reputation mechanism, for malicious nodes, the reputation is difficult to restore in short time, so RED of CBRM keeps comparatively stable. For KaZaa [14], RED is relatively low and frequently fluctuates, and that is mainly caused by lacking the security mechanism.

Figure 4 is the stability verification. In Fig. 4, AOT of CBRM is steadily increasing, owing to do some improvements such as considering the online time when electing the cluster head, and adding the virtual backup, etc. For KaZaa, because of just considering the limited factors when choosing super-node, so AOT of KaZaa is less than CBRM's.

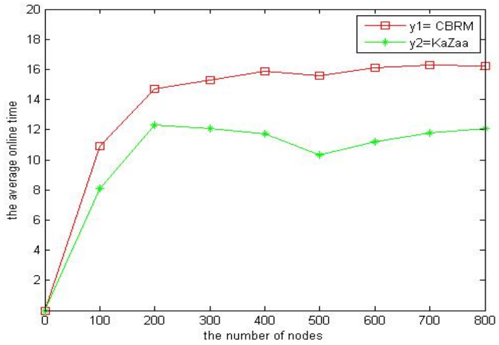

Figure 4. the average online time of all cluster head

Figure 5 is the test of network traffic. In the figure, the NT of CBRM is obviously better than KaZaa's. Because of improving on node join, node leave, disintegration, so after the formation of cluster, the cluster is comparatively stable. These improved measures reduce a lot of network flows. For KaZaa, it doesn't have these improvements, so network traffic is obviously high.

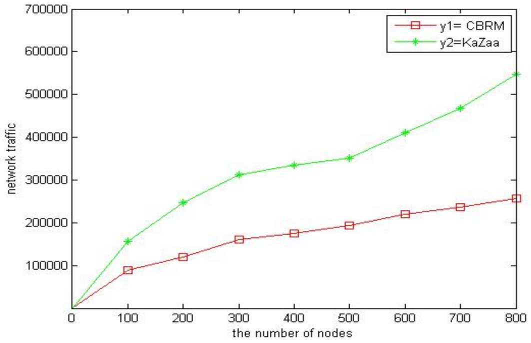

Figure 5. the network traffic

VI. Conclusion

For the security problems of P2P network, the paper proposed the cluster-based reputation model. The model uses the reputation mechanism to realize the security of transaction; it adopts cluster as the network topology; for the comprehensive performance of node, we select reputation, the historical average online time, the network bandwidth as the basic components; at the same time, the model improved on the cluster's initialization, node join, node leave, and the cluster's disintegration, and also posed the virtual backup. The results of simulation showed that the proposed model improved the system security, reduced the network traffic, and enhanced the stability of system.

References

[1] Li Bao, Application and Research of P2P Technology, computer development and application, Vol.22, pp.67-72,2009.

[2] JØsang A, Ismail R, Boyd C. A Survey of Trust and Reputation Systems for Online Service Provision[J].Decision Support Systems ,2005.

[3] P. Resnick, R. Zeckhauser, and E. Friedman, "Reputation Systems," Comm. ACM, vol. 43, pp. 45-48, Dec. 2000.

[4] Kamvar S D, Schlosser M T, Garciamolina H, The EigenTrust algorithm for reputation management in P2P networks[C] ,

Proeeedings of the 12th International worldwide Web Conference, BudaPest, 2003:640-651.

[5] Xiong L, Liu L, PeerTrust: Sopporting reputation-based trust for peer-to-peer electronic communities [J]. IEEE Transactions on Knowledge and Data Engineering, 2004, 16(7):843 一 857.

[6] Yao Wang, Julita Vassileva, "Trust and Reputation Model in Peer-to-Peer Networks,", Third International Conference on Peer-to-Peer Computing (P2P'03), pp.150, 2003.

[7] Beth T , Borcherding M , Klein B . Valuation of Trust in Open Network[C], Proceedings of the European Symposium on Research in Security, 1994.

[8] Claudiu D, Shahmehri N, Dynamic trust metrics for Peer-to-Peer system[C]. Proeeedings of the 16th Int1 Workshop on Database and Expert Systems Applications: 776-781, 2005.

[9] Damiani E, Capitani V S, Paraboschi S, etal, A reputation–based approach for choosing reliable resources in Peer-to-Peer networks[C], Proeeedings of the 9th Conferenee on Computer and Communieations Security:207-216, 2002.

[10] Mingxiao Hu, Jianli Li,Building a bybrid trust model for P2P systems, Computer Apllications, Vol.28 No.12, 2008.

[11] Li Jiang feng, Zhou Xingming, Zhang Zhenxi, Peer-to-Peer network with three tier topology based on anto clustering, Computer Science, Vol.36.No2,PP68.

[12] Royer E M, Melliar-Smith P M, Moser L E, An analysis of the optimum node density for Ad Hoc mobile networks[C], ICC. Helsinki: IEEE, pp:857-861, 2001.

[13] G D.Caro, F.Dueatelle, P.Heegaard, et al. Evaluation of basic serviees in ahn, P2P and grid networks [EB/OL]. http://www.es.unibo.it/bison/ deliverables /D07.Pdf, 2005.

[14] KaZaA file sharing network [EB / OL] . http : / / www . kazaa com/ ,2002.

Mei Chen received the Bachelor Degree in computer science & Technology from Tonglin University, Anhui, China, in 2008, He received the Master Degree in Computer Software & Theory from Donghua University, Shanghai, China, in 2011. Presently, he is just becoming the freshman doctor in University of Tokushima, in Japan. His present research interests include information retrieval, natural language processing.

Kenji Kita received the B.S. degree in mathematics and the Ph.D degree in electrical engineering, both from Waseda University, Tokyo, Japan, in 1981 and 1992, respectively. From 1983 to 1987, he worked for the Oki Electric Industry Co. Ltd., Tokyo, Japan. From 1987 to 1992, he was a researcher at ATR Interpreting Telephony Research Laboratories, Kyoto, Japan. Since 1992, he has been with the University of Tokushima, Tokushima, Japan, where he is currently a Professor in the Department of Information Science and Intelligent Systems. His current research interests include multimedia information retrieval, natural language processing, and music recognition. study should be lower-cased.

Xin Luo received the M.E. and D.E. degrees in Faculty of Engineering from The University of Tokushima, Tokushima, Japan, in 2004, and 2007 respectively. Since 2007, he has been with the School of Computer Science and Technology, Donghua University, Shanghai, China. His current research interests include natural language processing, and information retrieval.