The 5th International Conference on
Computer Science & Education
Hefei, China. August 24–27, 2010

Design and Realization of Distributed Nodes in a Multilevel-P2P Content Delivery Network

WANG Song, LING Qing, WU Gang, ZHENG Quan
Department of Automation of University of Science and Technology of China
Hefei,Anhui, 230027,China
wsong@ustc.edu.cn

***Abstract*—It is a great challenge to guarantee the service quality and reliability in the large scale peer-to-peer(P2P) networks of video-on-demand systems. This paper proposes a new multi-level P2P architecture, which can avoid the disadvantages of the centralized P2P and the pure P2P architectures while efficiently combine their merits. Methods are provided to design and realize the main distributed nodes. A prototype video-on-demand(VOD) system has been built. The experiments on that prototype system confirm that the expected functions of the main nodes have been achieved, and the designing and realization methods under the new multi-level P2P architecture can guarantee the system's quality of service, improve the system's reliability and extend the system's bandwidth.**

***Index Terms*—VOD,multi-level P2P,distributed nodes**

I. INTRODUCTION

It is shown in the reports of CNNIC that as of June 2009, there have been 338 million network users, among which there are 320 broadband network users and more than 200 million network video users[1]. The fast expansion of broadband network users is a great impetus to the broadband network services. As one of the most important network services, video-on-demand (VOD) service creates new business areas and is a bright spot in the information industry[2]. Media content distribution (MCD) is one key issue in the large scale VOD systems[3], [4], [5]. The peer-to-peer(P2P) technology can significantly improve the MCD performance and has caught much research interests[6]. It is reported that P2P applications generate 60 percent of the broadband network traffic during the day, 90 percent of the broadband network traffic during the night and 40 percent of the enterprise network traffic[1]. So the P2P technology is not only the source of broadband network customers, but also a very important network service supporter[7]. The main idea of the P2P technology is to change the centralized processing and storage into the distributed ones. It is shown in [6], [7] that the P2P architecture can greatly outperform the client/server(C/S) architecture in the MCD of VOD systems, e.g., resolve the bottleneck issues resulting from the competition of the single resource, avoid the service catastrophe due to a single node, optimize the system's performance, and significantly improve the service capability and quality.

[0]This work is supported by Chinese 863 Project (2008AA01Z147), Natural Science Research Program of the Anhui High Eduation Bureau(No. KJ2009A152) and National Key Supporting Technology Project (2008BAH28B04).

There are 3 typical architectures of P2P technology, including the centralized directory architecture, pure P2P network architecture and hybrid P2P network architecture [6], [7], [8].

(1) Under the centralized directory P2P architecture, the central server manages all nodes. This architecture has the advantages of simple maintenance, highly efficient searching and updating of shared resources. But it is vulnerable to single node failure, suffers from the poor reliability, poor security, and cannot be implemented in the large scale networks[9].

(2) The pure P2P network architecture does not require a central server[10], can efficiently tolerate the node failure and guarantee smooth performance even under the frequent node entering-exiting situation. Due to its flooding-based information searching strategy, this architecture occupies much network bandwidth, may easily result in network congestions and even network break-down, and cannot provide any performance guarantee[11].

(3) Under the hybrid P2P network architecture, all nodes are classified into ordinary nodes and super nodes[12]. These super nodes, which save the information of some ordinary nodes, make up a fast relay layer (all searches are done only among the super nodes). The super nodes, together with the ordinary nodes under their supervision, compose a multiple-layer system. The ordinary nodes not only provide the requested data to the super nodes, but also receive the needed data from the super nodes. This architecture has the advantages of good performance, great scalability and easy maintenance, and the disadvantages of the weak fault tolerability due to the strong dependence on the super nodes[13], [14].

In order to overcome the disadvantages of the aforementioned three P2P architectures, we propose a new multi-level P2P architecture for VOD systems, design and realize the distributed nodes of these systems. We have done some experiments on a testbed, which confirms the superiority of the architecture. Under the new multi-level P2P architecture, there are 5 types of nodes, including central media servers(CMS), directory servers(DS), media distribution centers(MDC), edge media nodes(EMS)and home media terminals(HMT).These 5 types of nodes compose a double-layer P2P network, which is comprised of the wide area P2P media content distribution network and the local area P2P media content distribution and service network. The present paper proposes some methods to design different types of nodes and confirms the good performance of the new architecture with some experiment data.

II. Introduction to the multi-level P2P architecture and its superiority

An VOD system with the multi-level P2P architecture is shown in Fig. 1. This system is comprised of the front-end operation supporting network and the media content distribution and service network. The front-end operation supporting network provides the necessary services for the system's operations, including media processing center(MPC), domain name server(DNS), service portal(SP), Digital copyright management(DRMS), operation supporting system(BOSS) and so on. The media content distribution and service network under the multi-level P2P architecture is the core of the VOD system, provides the services of deploying, storing and sharing the media data, and consists of the metropolitan area media content distribution network and the local area media content distribution and service network in Fig. 1. In the media content distribution and service networks, there are five types of P2P nodes, including the critical media server(CMS), the data server (DS), the media distribution center(MDC), the edge media server (EMS) and the home media terminal (HMT), which comprise the following double-layer P2P network.

(1) The metropolitan area media content distribution network(MANP2P) . It is at the first level to provide the media content storing and distribution service, and is made up of CMS, DS, MDC and EMS. The functions of these fundamental blocks are briefly introduced here. CMS store all the original media content data supplied by MPC in Fig. 1, and deliver the data directly or through MDC to EMS. MDC works as a data buffer, and enhances the CDSN's capability of deploying, storing and distributing data. EMS acts as the data node to store and distribute media content data.

(2) The local area P2P content distribution and service network(LANP2P), which is composed of EMS and HMT and lies at the second level of media content storing and distribution. EMS provides VOD and directory service to HMTs, deploys and distributes media content data in the LANP2P. HMT is the media terminal to provide users with video-on-demand service while functioning as a relay node to store and deliver data.

Under the traditional P2P architecture, nodes are loosely connected and may behave unpredictably, quality of service(QoS) of systems cannot be guaranteed and users cannot get the genuine VOD service. The centralized P2P architecture suffers from the fragility resulting from the single node failure, weak scalability and low reliability. Our new multi-level P2P architecture can efficiently combine the advantages of its two counterparts, provide QoS guarantee, and enhance the service capacity in the large scale systems. The major advantages of our multi-level P2P architecture is listed as follows.

Our new multi-level P2P system outperforms its conventional C/S VOD counterpart, which may be broken down by a single node failure and cannot guarantee QoS in the large scale because of its limited service capacity. Our system is also better than the traditional P2P one, which stores data in a lot of scattered nodes and may not provide any guarantee to QoS due to the loose node structure and the non-deterministic behavior of users. In our multi-level P2P architecture, the central control of the C/S VoD system and the distributed data sharing of the conventional P2P system is efficiently integrated together to avoid the bottleneck issues of servers, guarantee QoS, enhance the service capacity in the large scale configuration. The contributions of our system can be listed as follows.

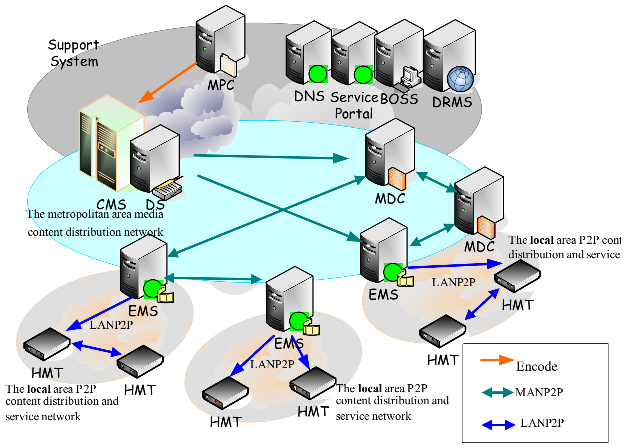

Fig. 1. Architecture of multi-level P2P VOD system

(1) The multi-level P2P structure. (i). The top level MANP2P stores and distributes data in the wide area network, and establishes efficient data sharing at that level by P2P technology. The high bandwidth of the back-bone networks and the optimal data deploying algorithms can significantly improve the efficiency of data storing and distribution in the MANP2P. (ii). The bottom level LANP2P stores and distributes data in the local area network based on P2P technology. In the LANP2P, EMS supplies the required media data to HMT, and HMT provides the VOD service to the final users. Besides the terminal functionality, HMT can also save media data automatically or according to commands of EMS, and supplies data to other HMTs. It is both EMS and HMT that can supply data in the LANP2P. Such abundant data resources can greatly increase service capacity, guarantee the video-playing quality, and ensure good scalability, which is confirmed by the experiments in the later prototype system.

(2) EMS plays a role of a super node. In the MANP2P, EMS works as a data node, stores and shares data by the P2P technology. In the LANP2P, EMS provides data service, responds to directory inquiry, actively deploys data according to the history of users' behavior and the prediction of hot videos, and monitors the system status. It is EMS that avoids non-determinism of nodes in the conventional P2P systems, and guarantees QoS.

(3) HMT can store and distribute data. Specifically HMT stores the history data according to certain criteria or commends from EMS. Furthermore HMT can supply the locally

stored media data to other HMTs while providing VOD service to its own users. Because of the functionality of HMT and the structure of the LANP2P, the burden of EMS is significantly reduced, the common video halting and accident breaking-down of conventional VOD systems can be effectively avoided, the system's scalability is greatly improved, and QoS in the large scale environment can be guaranteed.

III. Design and Realization of Distributed Nodes in the VOD Systems

A. Functional blocks of distributed nodes in the multi-level P2P architecture

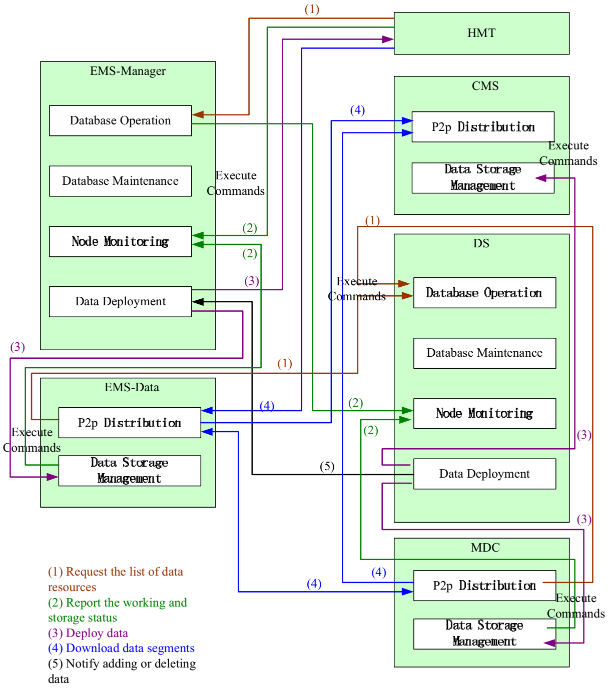

Fig. 2. Module structure of node

The typical functional blocks of distributed nodes are shown in Fig. 2. These blocks are designed to accomplish the important tasks, such as signal processing, database operation, database maintenance, node monitoring, data deployment, P2P distribution, data storage management. Now we briefly introduce some critical functional blocks. (1) The signal processing block processes datagrams, communicates with other blocks, manages the receiving, sending, encapsulation and decapsulation of datagrams. (2) The database operation block waits for the data query from other blocks, fulfills the requests, returns back the searching results (or updates the database), and generates and updates the list of sources of data. (3) The database maintenance block's main task is to periodically maintain the database, including data analysis and data backup, etc. (4) The node monitoring block is responsible for monitoring the status of nodes and the saved data, and timely updating the database operation block any change. (5) The data deployment block, at the requests of adding or deleting video from DS, deploys the media data among the nodes of the local area networks based on the status list of nodes. (6) The P2P distribution block provides media data uploading and downloading in the P2P way. (7) The data storage management block updates (adds or deletes) the media data saved on the nodes by some criteria.

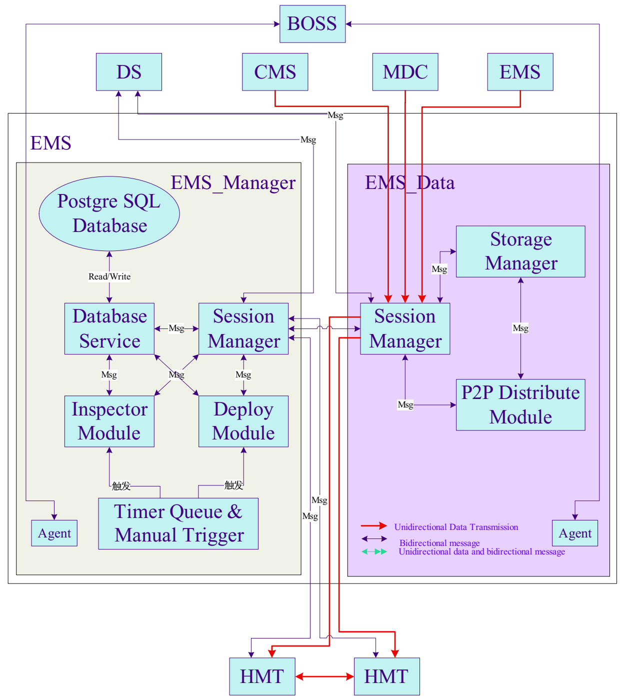

Fig. 3. Structure of EMS

B. The edge media server (EMS)

After inquiring the directory server(DS), EMS locates the target videos, downloads the relevant data from CMS or MDC by the P2P technology, and satisfies the media data requests of HMTs (the users can watch a video after having completely downloaded it or while downloading it). The major functions of EMS has three major functions, including

(1) Media data transmission. EMS can receive data from CMS, MDC, HMT and other EMS'. It can also transmit data to HMTs and other EMS'.

(2) VOD service. By the P2P technology, EMS can provide HMTs the requested video service.

(3) Data deployment. EMS efficiently deploys media data in the LANP2P so that the network traffic is well balanced and QoS of the VOD system improves.

EMS is comprised of two parts, EMS-Manager and EMS-Data as shown in Fig. 3. EMS-Manager provides directory service, deploys data, monitors the system status, maintains the database, etc. EMS-Data is responsible for storing and distributing data. One EMS-Manager can supervise multiple EMS-Data. EMS-Data can be added into the system in the Plug & Play fashion. Now we briefly go through some major functional blocks of EMS-Manager. "Postgre SQL Database" keeps the directory information. "Database Service" reads the concerned data of "Postgre SQL Database" into the high speed cache so that the inquiry requests from other nodes can be timely responded. "Inspector Module" is in charge of the inspection of the LANP2P. Data deployment in the LANP2P is the major function of "Deploy Module", which is driven by either the events generated by timers or manual intervention. EMS-Data is equipped with an intelligent "Storage Manager", which updates "Postgre SQL Database" with all modifications of the locally stored data. "P2P Distribute Module" in EMS-Data manages data downloading and uploading (more specifically, setting the bit rate of data transmission and controlling the number of downloading and uploading paths), and can supply data not only to nodes in the WANP2P but also HMTs in the LANP2P. When the data requested by a HMT is not available, "P2P Distribute Module" retrieves that data from other nodes in the WANP2P and relay it to the target HMT. Most blocks of EMS are message-driven. "Session Manager" receives and sends all messages in EMS, interprets the messages by the pre-defined rules, and is well scalable and compatible.

The functional blocks of EMS are designed in a block-structured way, and have the advantages of error isolation, easy maintenance, consistent interface, good adaptivity and certain fault tolerance.

C. Directory server (DS)

DS monitors the entrance and exit of nodes in the metropolitan area network and provides the directory service to the nodes in MANP2P. Moreover DS is responsible for database synchronization with EMS so that the data in the MANP2P is consistent with that in the LANP2P. DS also analyzes the users' VOD behavior to realize the balanced deployment of the media data in the VOD systems.

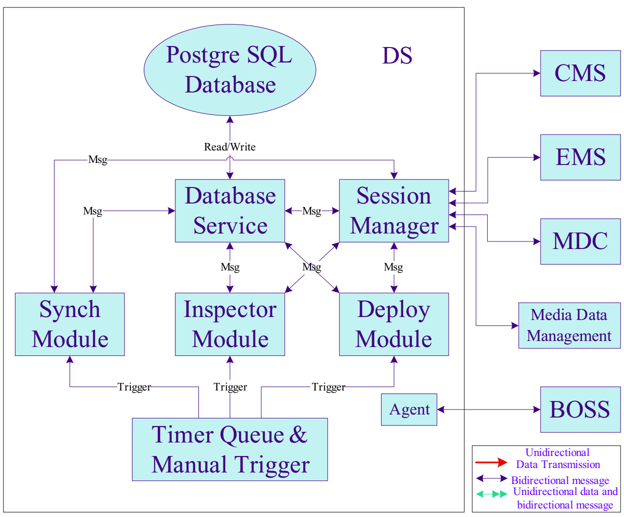

Fig. 4. Structure of DS

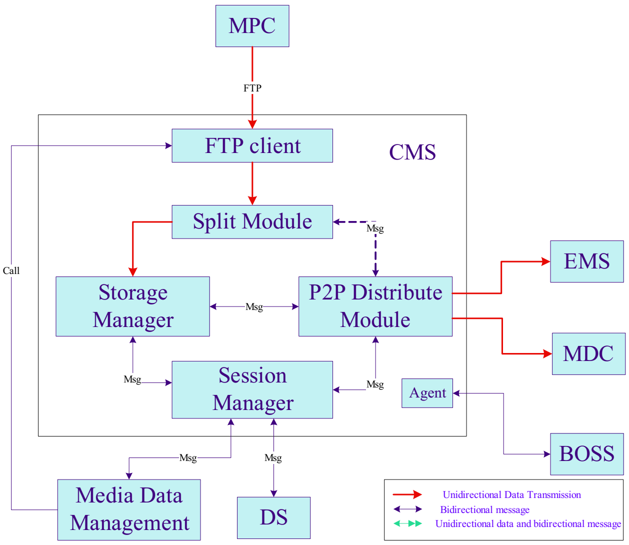

Fig. 5. Structure of CMS

The structure of DS is shown in Fig. 4. The "Postgre SQL Database" block maintains the directory information. The "Database Service" block executes the database related operations, responds to the inquery from other nodes and prioritizes the changes on the database. DS can be manually or periodically triggered to do synchronization, monitoring and deployment. The software modules of DS are realized in a block-structured way so that they can work independently, is scalable and easy to maintain, and can tolerate some faults.

D. Central media server (CMS)

CMS is responsible to distribute data to nodes in the MANP2P. CMS downloads coded and encrypted media data from MPC by the FTP technology, processes and packetizes the obtained data, and updates DS the data information.

The fundamental structure of CMS is shown in Fig. 5. FTP client block gets the encrypted media data and the relevant XML description files from the shared data servers. It locally packetizes the received media data. A data packet consists of 15 MB, is saved in a directory structure. The HASH values of the source data and the data packets are also saved for fast search. Storage Manager block intelligently coordinates the structural storage and release of media data on the CMS, and reports to DS the data status so that CMS is synchronized with DS regarding the media data and the integrity of data can be guaranteed. P2P Distribute Module block uploads data, dynamically adjusts uploading parameters (including the bit rate, the number of transmission threads) to guarantee QoS, analyzes the behavior of node and reports DS the analysis results.

E. Media distribution cache (MDC)

MDC is the data node in the MANP2P. Media data are saved on MDCs in a distributed format so that the reliability and stability of the systems is improved. The structure of MDS is shown in Fig. 6. Storage Manager intelligently manages the storage of media data to guarantee that its data is consistent with that in the database. P2P Distribute Module is in charge of uploading or downloading data, and realizing the media data distribution in the MANP2P, i.e., providing data to other nodes in the MANP2P.

IV. A PROTOTYPE SYSTEM AND THE EXPERIMENTAL RESULTS

Fig. 6 shows a prototype system which is built in Shanghai Jiaotong University. CMS and DS are installed on the same server in Lab 1. EMS1 and MDC1 are placed in Lab2, and are realized by two separate servers. EMS2 and MDC2 are two separate servers in the computer lab of the Minhang Campus. There are hundreds of HMTs, which are located in the labs, computer rooms and student dorms, and provide VOD service to users. In this prototype system, we did experiments under two different schemes. One does not implement P2P technology; the other implements the proposed new multi-level P2P technology. The experimental results of the two schemes are shown in Table 1.

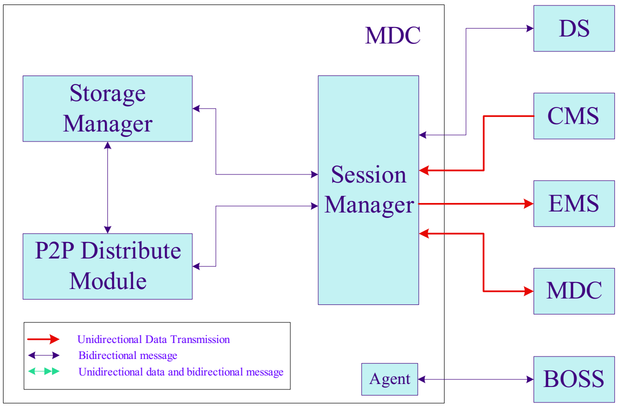

Fig. 6. Structure of MDC

It can be seen in Table 1 that the multi-level P2P architecture significantly improves the capacity of the VOD system, greatly reduces the burdens of servers, efficiently extends the system's bandwidth, and can guarantee the system's QoS. In our prototype system, only two ordinary EMSs can support more than 1000 concurrent VOD threads and serve more than 200 HMTs. Such performance is by far better than that without P2P technology.

V. CONCLUSION

The wide-spreading applications of broadband networks has laid a solid foundation for the VOD service, which is growing into an important profitable industry in the near future and undergoing fast growth. This paper proposes a new multi-level P2P architecture, which can efficiently utilize the bandwidth of the idle network edge nodes, significantly reduce the traffic burdens on the back-bone nodes of the network, greatly lower

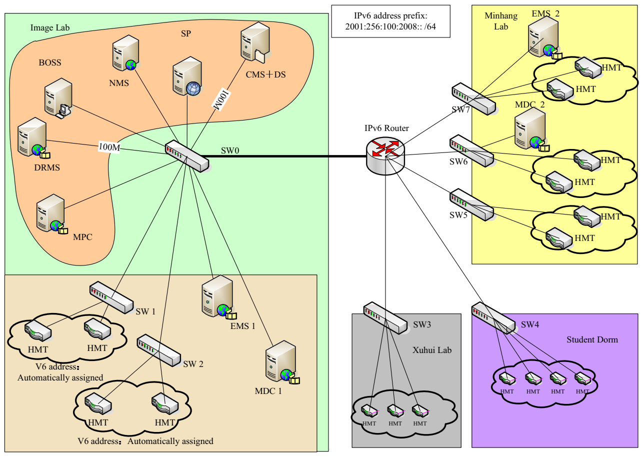

Fig. 7. Topological structure of testing system

TABLE I
TABLE 1 EXPERIMENT RESULTS

	No P2P	P2P
TCP Data of EMS	11.56GB	3.65GB
UDP Data of EMS	0.14GB	0.15GB
Total Data of EMS	11.7GB	3.8GB
Total Data of HMTs	0	7.9GB
Total user Data	11.7GB	11.7GB
Video quality	Poor	Good

the requirements on the servers and the user terminals so that the system's cost is significantly reduced. The new architecture is very scalable and has great potentials in the future large-scal VOD systems.

The design and realization of distributed nodes is the bottle-neck issue in the multi-level P2P VOD system. This paper provides some methods in this area and a prototype system is built based on these methods. The experimental results on the prototype system confirms the efficiency of the proposed design and realization methods. To further improve the reliability and stability of distributed node will be our future work, which will optimize the system's structure and upgrade the overall system performance.

REFERENCES

[1] *25th China Internet Development Statistics Report*,CNNIC,2010-1-15.http://www.cnnic.cn/html/Dir/2010/01/15/5767.htm.(in Chinese)
[2] Abdelhamid Nafaa,Sean Murphy and Liam Murphy. Analysis of a Large-Scale VOD Architecture for Broadband OperatorsA P2P-Based Solution,*IEEE Communications Magazine*,2008,46(12):47-55.
[3] Yang Ge,Liao Jian-xin,Zhu Xiao-ming et al. Survey of Key Technologies of the Distribution System for Streaming Media[J],*Acta Electronica Sinica*,2009,37(1):137-145(in Chinese).
[4] Liu Yu,Zhu Guang-xi,Yin Hao et al. Content Delivery/Distribution Networks :Overview[j]. *Computer Science*,2009,36(2):12-15(in Chinese).
[5] Jiang Hai,Li Jun,Li Zhong-cheng. Hybrid Content Distribution Network and Its Performance Modeling,*Chinese Journal of Computers*,2009,32(3):474-482(in Chinese).
[6] Peng Zhao,Yang Zong-kai,Chen Jing-wen et al. Design and Evaluation of a Scalable P2P Video on Demand System[J],*Journal of Chinese Computer Systems*,2009,30(9):1730-1733(in Chinese).

[7] Peng Kai,Wu Juan,Yang Zong-kai et al. Research and Prospect on P2P Live Streaming System[J]. *Computer Science*,2009,36(1):10-15(in Chinese).
[8] WANG Xiao-ming,YAO Zhong-mei,Dmitri Loguinov,Residual-Based Estimation of Peer and Link Lifetimes in P2P Networks,*IEEE/ACM Trans. Networking*, 2009,17(3):726-739.
[9] YAO Zhong-Mei,WANG Xiao-Ming,Derek Leonard. Node Isolation Model and Age-Based Neighbor Selection in Unstructured P2P Networks[J],*IEEE/ACM Trans. Networking*,2009,17(1): 144-157.
[10] Gu Ke,Wei Reng-yong. Improved Research of P2PModel of CAN[J],*Journal of Chinese Computer Systems*,2007,28(3):416-420(in Chinese).
[11] Fabin E. Bustamante and QIAO Yi. Designing Less-Structured P2P Systems for the Expected High Churn[J], *IEEE/ACM Trans. Networking* ,2008,16(3):617-627.
[12] Guo Liang-ming, Yang Shou-bao,Guo Lei-tao, et al. Supernode Selection Mechanism Based on District Partition in Peer- to-Peer Networks[J] *Journal of Chinese Computer Systems*, 2008,29(2):208-212(in Chinese).
[13] Xia Qi-zhi, Xie Gao-gang, Min Ying-hua. IS2P2P : Index Based Structured P2P Networks[J]. *Chinese Journal of Computers*,2006,29(4):602-610(in Chinese).
[14] Duc A Tran,Kien A. Hua and Tai T Do. A Peer-to-Peer Architecture for Media Streaming[J],*Journal of Selected Areas in Communications*,2004,1-14.