© 20xx IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.
The final published version is available at http://dx.doi.org/10.1109/SSST.2005.1460891.

Peer-to-Peer Based Multimedia Digital Library

Yi Ma **Ramazan. S. Aygün**
Computer Science Department
University of Alabama in Huntsville
yma@cs.uah.edu, raygun@cs.uah.edu

Abstract: *New developments in super peer based Peer-to-Peer network and digital library techniques give us the inspiration of looking for an approach to combine the reliability and strength of server with the flexibility of P2P network in multimedia data sharing. In this paper we describe the design issues of a Peer-to-Peer based multimedia digital library. Its design is based on super peer based P2P network using HyperCuP topology. Peers build local digital library. They publish their local data in network and interoperate with other peers through Open Archive Initiative (OAI) interface. I-Super-Peers process user requests and find all available data. G-Super-Peers help requesting peer to generate distributed caching and streaming plan for user data request. The plan should be an optimized one such that it will increase overall performance.*

Keywords: Peer-to-Peer System, Multimedia Digital Library, Multimedia Streaming.

1. Introduction

Peer-to-Peer (P2P) systems gain much attention in recent years. It is considered as a new distributed computing model and is a competitive alternative for Client/Server (C/S) model. On the other hand, World Wide Web (WWW) has dominated the Internet and accommodates billions of Web pages. Many organizations, companies or individuals communicate and share their data with others over the Internet through various services provided by servers. Since such a huge environment has been well developed and seriously depends on the C/S model, P2P could not be expected to replace C/S model at least in near future. The other reason that supports the survival of C/S is that many services like digital library require servers to support acceptable performance. However, the high requirements for being a server significantly limit Internet users' ability to provide their own data and service. The pure consumer form of client also wastes considerable resources such as idle CPU cycles, bandwidth and storage spaces. This in turn aggravates servers' working load. It is quite reasonable to think about finding some way to combine them together such that we could gain the advantages of both as well as avoid their weakness. However, there is still much work before we could reach a point such that the strengths of two are well balanced.

Early P2P systems are weak in sharing information due to their lack of ability to represent semantics. In some of recent projects like Edutella [1]and DBGlobe [2], researchers begin to delve efficient semantics handling of shared content by schema based P2P network in which topology of overlay network is combined with information schemas. However, these systems most contribute to text based data sharing. When consider sharing multimedia data, especially as to multimedia streaming, they are not intuitively suitable. In projects like PROMISE [3]and Zigzag [5], they distribute steaming tasks in P2P network through topology aware approaches. However, these projects either impose too strict constraints over network topology or have logical error in streaming strategy. The other main weakness of these systems is that they didn't provide mechanisms for information publishing, discovery and replication.

As an important information sharing technique, Digital Library (DL) provides systematic mechanisms for document management, information retrieval, and data delivery. Traditionally, DL systems usually contain huge volume of digitized document and use reliable equipments to support their powerful services. Furthermore the recent Open Archive Initiative (OAI) framework [9] [8], shown in Figure 1, based on HTML, XML, and Dublin Core, provides easy-to-implement protocols to achieve interoperability among different DL systems. This enables individual user to build their own local DL interoperable with others.

Under such a scenario, we propose a P2P based multimedia digital library network that handles document publishing, storage and replication management, information retrieval, and streaming for digitized multimedia data.

In the section 2 we will describe the application scenario of P2P based digital library.

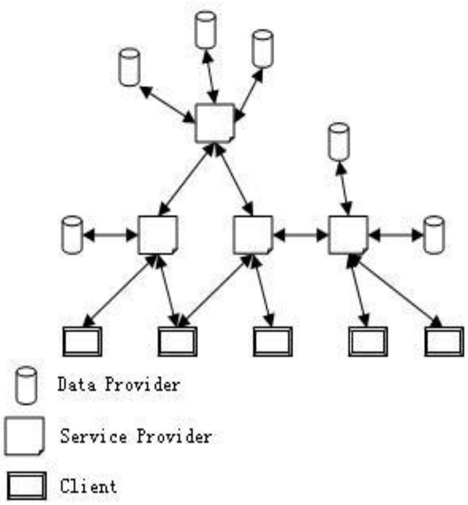

Figure 1: OAI Topology

2 Application Scenario of P2P Multimedia Digital Library

New developments in super peer based P2P network and digital library techniques give us the inspiration of looking for an approach to combine the reliability and strength of server and the flexibility of P2P network in multimedia sharing.

As digital library systems accommodate huge volume video/audio data accessible to Internet user, individuals also create their own films record funny or miraculous things happened in their daily life. They may be ebullient to share their masterpieces with others. Our P2P based multimedia digital library endeavor to build a flexible environment to facilitate automatic document publishing and information retrieval such that storage, replication and streaming task are well distributed over the P2P infrastructure. Hence the overall capacity, including things like storage space and outbound bandwidth, and reduce the working load of single resource.

The underlying P2P system is a Super Peer based network. Super peers maintain the metadata information and are organized as communities around semantically related ontology. This is implemented as a HyperCuP [6] based network topology. There is another form of super peer, which maintains bandwidth information of geographically aggregated peers to facilitate distributed data buffering and streaming. Individual peers extract metadata of local schema from their data and setup local knowledge repository. Then they register their local digital library or publish their new document, through sending OAI-identify statements that contain above metadata. These statements contain the target spaces of query request and respondent. These statements are delivered to appropriate super peer. The super peer add metadata contained in identify statements into their repository. These super peers are called I-Super-Peer. Furthermore, whenever a peer is connected to the network, they register sizes of their contributable buffer and bandwidth to P2P network. These resources will be harnessed to facilitate data caching and streaming distribution. This information is directed to geographically nearest super peer. These super peers contain bandwidth and storage space information and are called G-Super-Peer. A super peer may be both I-Super-Peer and G-Super-Peer. Use a server to work as super peers will help the reliability and availability of the P2P network.

The process of online reviewing requests of some multimedia data is implemented by two independent transactions. Firstly, a peer delivers its queries to I-Super-Peers to find intended data. The I-Super-Peers process query within super peer backbone. I-Super-Peers are organized as HyperCuP and form ontology communities. They will return requesting peer all available data. Then the peer selects a most appropriate document and send streaming request to G-Super-Peers. These G-Super-Peers process streaming request within G-Super-Peer backbone and help requesting peer to generate caching plan and streaming plan. The streaming plan is returned to requesting peer. The caching plan is also sent to those peers who contain requested data source as well as those peers who are ready to contribute their buffer space and bandwidth.

In above description, we assume that there exists a naming and locating mechanism. That is, a unique location independent identification could be generated to identify each object exists in the system. Furthermore, we could locate correct object through identifying its name and deliver our data.

3 Architecture and Services of P2P Multimedia Digital Library

In following Figure 2, we show a simple framework of our P2P based Multimedia Digital Library.

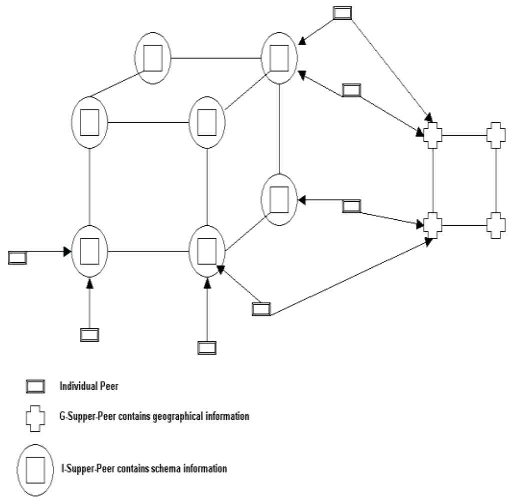

Figure 2: P2P Multimedia DL Topology

Basically, there exist three peer types. The individual peers are normal peers that organize their local document as a local digital library. They will contribute their data, storage/buffer spaces, and bandwidth to the P2P system.

I-Super-Peers contains schema metadata about specific ontology. They are organized as the HyperCuP backbone. They are service providers[9] that "harvest" schema metadata from joined peers. A variation of these service providers is that they permit peers positively register their metadata. The schema gathering is implemented through OAI interface.

G-Super-Peers are organized as a separate HyperCuP backbone based on geographical their characteristics. For example, all G-Super-Peers in Madison county of Alabama may organize as a cube. This cube will further join the cube of G-Super-Peers in state of Alabama. Whenever a peer is online, it should register its contributable cache space and bandwidth space information to nearest G-Super-Peer. G-Super-Peers use this information to generate caching plan and streaming plan for specific streaming request. A set of peers connected to same G-Super-Peer form a Geographical Community.

The P2P multimedia digital library will at least provide following services: document publishing, data replication, information retrieval, and distributed cache and streaming scheduling.

Document publishing is the process of register local data to P2P network. Initiative peers are responsible to provide appropriate schema metadata for their document. The registration is implemented through OAI interface.

Data replication service is the process to replicate frequently accessed data to where the access is generated. Data replication is implemented by G-Super-Peers under the authority of data resource. Data resource may deny the replication and is responsible to inform G-Super-Peer the updating information.

Information retrieval is the process to find requested data. This service is implemented by I-Super-Peers through schema information.

Distributed cache and streaming scheduling will be discussed in following section.

4 Caching and Streaming in P2P Multimedia Digital Library

After the appropriate data has been found, the most important thing is to find appropriate resources and plan to perform the streaming task. We should also endeavor to design efficient replication plan to improve future streaming tasks.

In the P2P environment, there are no restrictions on when and where a peer could join or leave the system. However, as to multimedia streaming, the location and time when the peer is available are critical. So in our system, any individual peer should register to nearest G-Super-Peer when they are connected. Furthermore, each peer has the responsibility to contribute part or its storage space and bandwidth to the system. Other peers may harness these resources on their need.

Firstly, a requesting peer sends its query to all connected I-Super-Peers. These I-Super-Peers contain schema information about all the peers in the system, whatever the peers are connected or not. User query are processed within the I-Super-Peer backbone and all data resources are found and returned to requesting peer. This is feasible as to its implementation in Edutella.

When the requesting peer received the information of data resources, it is the user's responsibility to remove overlap information and choose most wanted document. Remind that a single document may exist in multiple places and these places will all be found by I-Super-Peers.

Then the user sends information about chosen document and related data source information to G-Super-Peers. G-Super-Peer finds all available idle buffer and bandwidth resources that are geographically near the requesting peer, i.e. peers with small hops from requesting peer and contribute sufficient buffer and bandwidth. Although the G-Super-Peer would not directly generate caching and streaming plan for requesting peer, it will balance the load among all data requests.

The information about data sources and idle resources is returned to requesting peer. The requesting peer is responsible for generating efficient caching and streaming plan and sends back plan information back to G-Super-Peer. Following figures shows the framework of caching and streaming strategy of our P2P multimedia digital library.

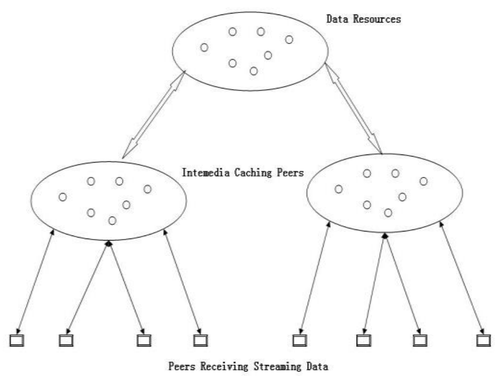

Figure 3: Framework for Caching and Streaming

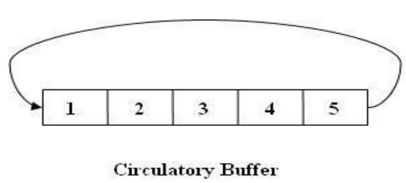

Figure 4: Circulatory Buffer Structure of Receiver

A caching plan departs the source data into several segments and assigns each part to an intermediate buffer. These intermediate buffers are far nearer to receiver that the data sources. Considering there are may exist multiple peers possess requested data. Data segments are transfer from source peer to most appropriate intermediate buffer based on bandwidth distribution. The streaming plan assigns each chosen intermediate buffer peer a streaming task with an approximated time constraint. This constraint defines the latest time when the buffer should have already prepared assigned data segment.

The receiver streams data from these intermediate buffers. To fully utilize the idle resources in the P2P network, the receiver, i.e. requesting peer, use a circulatory buffer. The buffer is divided into several parts. The division may be either static, which is based on arbitrary choice of receiver, or dynamic, which is based on the size of document, available intermediate buffer and bandwidth. The receiver will decode buffered data sequentially. After a segment of buffer is fully decoded, the receiver will perform two tasks. Firstly, it goes on decoding following segment. At same time it may either discard the decoded part or put them into disk storage space. Secondly, the receiver will refill the just decoded buffer segment by streaming new data from appropriate intermediate buffer.

If there are too many or frequent requests for same document from single geographical community, the G-Super-Peer is responsible to detect this situation and replicate the document in its community. These replications are used for future requests. This approach will reduce the overall working load of P2P network and increase streaming efficient within the community. When requests of the document has reduced under a thresh hole value, replications are deleted from the community.

The algorithm used for generating optimized caching and streaming plan could be implemented by finding an optimal power assignment of the graph consist intermediate buffers and corresponding network paths.

5 Conclusions

In our P2P based multimedia digital library system, we defined a complete set of mechanisms enable the management and delivery of multimedia data in interconnected data repositories. This system is expected to provide high performance for P2P multimedia streaming.

Reference:

[1] W. Nejdl, B. Wolf, C. Qu, S. Decker, M. Sintek, A. Naeve, M. Nilsson, M. Palmr and T. Risch *EDUTELLA: a P2P Networking Infrastructure based on RDF.* 11th International World Wide Web Conference Hawaii, USA, May 2002

[2] E. Pitoura, S. Abiteboul, D. Pfoser, G. Samaras, M. Vazirgiannis, et al, *DBGlobe: a service-oriented P2P system for global computing*, SIGMOD Record 32(3): 77-82 (2003)

[3] Mohamed Hefeeda, Ahsan Habib,Boyan Botev, Dongyan Xu, Bharat Bhargava . *PROMISE: Peer-to-Peer Media Streaming Using CollectCast.* MM'03, November 2–8, 2003, Berkeley, California, USA.

[4] Dongyan Xu, Mohamed Hefeeda,Susanne Hambrusch, Bharat Bhargava. *On Peer-to-Peer Media Streaming.* ICDCS archive Proceedings of the 22 nd International Conference on Distributed Computing Systems (ICDCS'02) Page: 363

[5] D. A. Tran, K. Hua, and T. Do, "*A peer-to-peer architecture for media streaming*," IEEE Journal on Selected Areas in Communications, Special Issue on Advances in Service Overlay Networks, vol. 22, pp. 121--133, Jan 2004.

[6] M. Schlosser, M. Sintek, S. Decker, and W. Nejdl. HyperCuP---Hypercubes, Ontologies and Efficient Search on P2P Networks. In International Workshop on Agents and Peer-to-Peer Computing, Bologna, Italy, July 2002.

[7] Karnstedt, M.Hose, K.Sattler,K.-U.TU Ilmenau; *Query routing and processing in schema-based P2P systems*. Database and Expert Systems Applications, 2004. Proceedings. 15th International Workshop on 30 Aug. - 3 Sept. 2004 page(s): 544- 548

[8] Ahlborn, W. Nejdl, and W. Siberski. *OAI-P2P: A Peer-to-Peer Network for Open Archives.* In Workshop on Distributed Computing Architectures for Digital Libraries - ICPP2002, 2002.

[9] C. Lagoze, and H. Van de Sompel, *The open archives initiative: building a low-barrier interoperability framework*. ACM/IEEE Joint Conference on Digital Libraries, 2001.