12th Optoelectronics and Communications Conference/16th International Conference on Integrated Optics and Optical Fiber Communication (OECC/IOOC2007)
Technical Digest, July 2007, Pacifico Yokohama

13A2-5

Prefetching Protocol Proxy with Optimal Mirror Selection and Burst Transmission

Tomohiro TSUJI, Junichiro HONMA, Sho SHIMIZU, Yutaka ARAKAWA, and Naoaki YAMANAKA
Department of Information and Computer Science, Faculty of Science and Technology
Keio University, 3-14-1 Hiyoshi, Kohoku-ku, Yokohama, 223-8522, Japan
Email: tsuji@yamanaka.ics.keio.ac.jp

***Abstract*— In this paper, we propose a new accelerated download mechanism for huge data such as rich contents that can reduce the download time by about seventy percent as compared with the conventional manual operation.**

***Index Terms*— prefetching, mirror server selection, protocol conversion**

I. Introduction

The rich contents such as movies and music transmitted over the Internet have been increasing due to the explosive growth of Web services and the bandwidth of network. Also, larger files are exchanged over the Internet with Peer to Peer application. Then, the flash crowd problem have surfaced as the number of users who access to the Internet [1]. When file size is large, it is difficult to predict the number of accesses in flash crowd from the number of accesses in ordinary times. Therefore, many load-balancing methods such to deliver the large files to a lot of people have been proposed.

In this paper, we propose a new accelerated download mechanism for large data stored in multiple mirrors. Our proposed scheme includes prefetching, automatic optimal mirror selection, and protocol conversion. In the proposed scheme, we measure the response time from the client to all mirror servers. This enables us to save the time to download contents. However considering the delivery of HDTV movie contents, the file size is equivalent to several GBytes. Therefore, the download time increases even if such scheme is used. Then, we apply prefetching technology to this problem. By applying prefetching, it can reduce the feeling download time. Moreover, we propose protocol conversion technology in order to improve the real download time.

In this research, we mount these systems as a proxy and evaluate their performance. Implementation result shows the proposed scheme can reduce the download time by about seventy percent as compared with the conventional manual operation.

II. Conventional Scheme

Several mirror server selection schemes have been proposed for delivering the large files to a lot of people. The First one is server initiation type such as DNS (Domain Name Server) Filter and Round Robin [2] where requests from client are distributed automatically at server side. The Second is client initiation type where mirror server list is given to the client and the client selects the optimal mirror server. The former is chiefly used for the Web server and the streaming server, and has the advantage of transparency, i.e. the client need not consider the mirror server. But the client's request is not always forwarded to the nearest mirror server. And also when the domain is different, this method cannot be used. While, the latter is used to deliver the software in Sorceforge.net. If there are a lot of mirror servers, the links to each mirror server with country information are often shown to client. The client have to select the best mirror server based on past experience and knowledge, but the client's request is not always forwarded to the nearest mirror server.

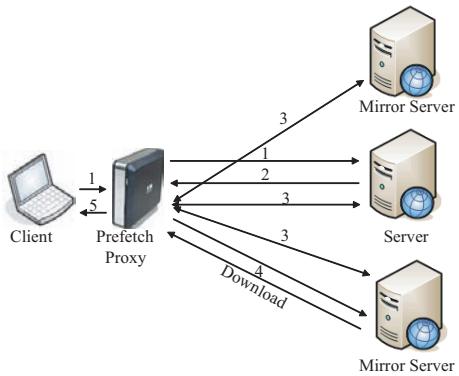

Fig. 1. Prefetching Procedure of Proposed Scheme.

III. Proposed Scheme

In our proposed system the clients are connected to the contents server via prefetching proxy server. And there are a lot of mirror servers which store large files. Our proposed system is composed of prefetching, automatic optimal mirror selection, and protocol conversion. We show that our proposed scheme can greatly reduce the download time.

A. Mirror Selection Scheme & Prefetching

We propose an automatic optimal mirror selection scheme which selects the nearest mirror server among a lot of mirror servers. Figure 1 shows the procedure of our proposed scheme. First, a client accesses a web server through a prefetching proxy server (procedure 1) . Second, in response to client's access, a prefetching proxy server analyzes the HTML document received from Web server, and extracts the links to all mirror servers (procedure 2) . Third, a prefetching proxy server selects a download link with shortest response time (procedure 3) . Fourth, a prefetching proxy server requests content to the mirror server with shortest response time. A prefetching proxy server prefetches the content and stores in cache (procedure 4) . Fifth, if a client clicks the link, the content will be downloaded from prefetching proxy server instead of Web server (procedure 5) . Our system can shorten

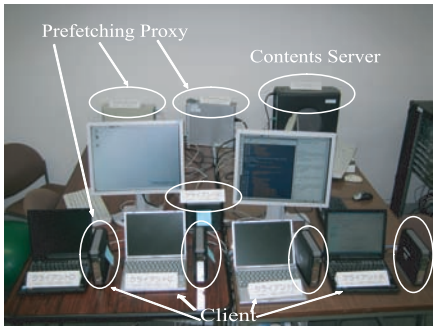

Fig. 2. Experimental system of our proposed scheme.

the real download time by selecting the optimal mirror server. And also, we can reduce the feeling download time by prefetching the content. Well then, how to select an optimal mirror server is an important issue in our system. We examine two methods such as ping and netselect.

1) Ping: Ping is a program that diagnoses TCP/IP networks such as the Internet and Intranet. In this research, we select the mirror server based on the web access time measured by ping packets. Ping packets are transmitted to all mirror servers after prefetching proxy server extracts the link of all mirror servers, and RTT are measured. After that, prefetching proxy server selects the optimal mirror server.

2) netselect: Netselect is a command to select the nearby mirror server. To use netselect, we have to give the list of mirror servers as the argument. After we set the list, the scores of each mirror server are calculated. Then the mirror server which has the shortest web access time is selected automatically. The calculating formula of the score is as follows. In this research, we employ netselect to select the optimal mirror server of rich contents. The scores are calculated by computing RTT, the number of hops and the download time.

$$tmp = \frac{Total\ RTT\ *\ Number\ of\ transmission\ packets}{Number\ of\ receiving\ packets} \quad (1)$$

$$score = tmp\ + \frac{(tmp\ *\ Number\ of\ minimum\ hops)}{10} \quad (2)$$

First, *tmp* is calculated by formula (1). Second, score is calculated by formula (2) using the result of formula (1). By the above formulas. Netselect selects the mirror server which have the minimum score [3].

B. Protocol conversion

Most of the Internet models is Server-Client model by HTTP. Because the request from the client to the contents server is very small signal, there is little possibility that a long delay is generated. But when the content size is very large, the delay increases. So, we propose the protocol conversion scheme to save the real download time. This indicates that HTTP requests are transparently alternated by other protocol such as P2P or FTP.

IV. Evaluation

We evaluate the mirror server selection time and the download time on our experimental system. Figure 2 shows

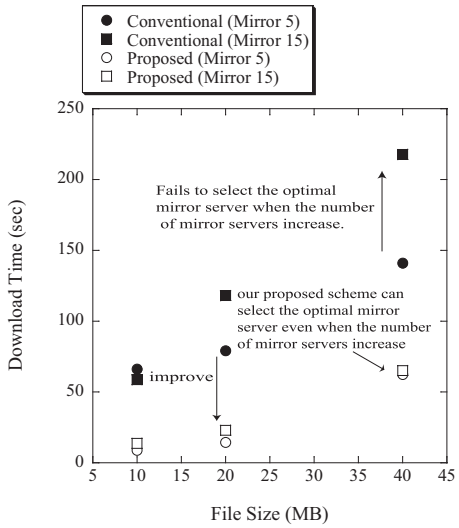

Fig. 3. The effect of Prefetching proxy.

the experimental system of our proposed scheme. There are seven clients connected with the prefetching proxy servers respectively. In this evaluation, since we need many large files distributed in a lot of world wide mirror servers, we use sourceforge.net as the contents server. The object comparison is the time that man downloads the file from the contents server described by HTML with a lot of mirror servers. After this, the total download time indicates the time which include the mirror server selection time and download time. Figure 3 shows the total download time versus the file size. We evaluate the total download time about the six cases. The performance of proposed scheme is greatly improved in total download time compared with the conventional scheme in all case. The conventional scheme falils to select the mirror server which has the shortest web access time when the number of mirror servers increase. However, our proposed scheme can select the optimal mirror server quickly even in that case and also prefetching the content. That is the reason we can reduce the total download time greatly compared with the conventional scheme. In our proposed scheme, it is possible to reduce the download time by about seventy percent as compared with the conventional scheme.

V. Conclusion

In this paper, we have proposed a new accelerated download mechanism for huge data such as rich contents that can reduce the download time by about seventy percent as compared with the conventional manual operation.

References

[1] Ratul Mahajan, Steven M. Bellovin, Sally Floyd, John Loannidis, Vern Paxson, and Scott Shenker,"Controlling High Bandwidth Aggregates in the Network," *ICSI Center for Internet Research (ICIR) AT&T Labs Research*, unpublished.

[2] John Dilley, Bruce Maggs, Jay Parikh, Harald Prokop, Ramesh Sitaraman, and Bill Weihl,"Globally Distributed Content Delivery," *IEEE INTERNET COMPUTING*, September・October 2002.

[3] APT HOWTO http://www.debian.org/doc/manuals/apt-howto/ch-basico.ja.html