A Proxy-based Architecture for Multimedia Transmission

WANG DA-ZHEN[1,2], WAN FANG[1]

(1.Department of Computer Science, Hubei University of Technology, Wuhan, P.R.C 430068)

(2.Department of Information Management, Wuhan University, Wuhan, P.R.C 430074)

http://www.newtelesoft.com

Abstract: Streaming media data have strict delay and loss requirements, which cannot be adequately supported by the current Internet. Proxy caching technology is introduced to alleviate the effect of the bottleneck bandwidth on the transporting quality. A new architecture of multimedia proxy caching is given to solve the problem of the traditional web caching technology to the multimedia streaming. The mixed media management module and the resource management module is added in this architecture, the prefetching and replacement method is studied considering the characteristics of streaming media data, the concept of weighed popularity is given to enhance the exactness of the replacement algorithm.

Keywords streaming media; proxy caching; quality of service; replacement algorithm

With the rapid development of the Internet and streaming media service on the Web, there appear many Internet streaming media applications, such as distance education, video conferencing, video-on-demand, E-Commerce [1], which become more and more popular. Streaming media is very large that requires more bandwidth, which will lead to significant increase in latency and network congestion. The Internet is best-effort networks and cannot give any quality of service guarantees to streaming media transmission. Therefore, there are many challenges on the transport of video streaming. This paper provides a solution to solve this problem using multimedia proxy caching [2]. A proxy cache resides close to the clients, handles the clients' request, if the request data is in the cache, it returns the data directly to the client, if not, it forward the request to the right original server or other proxy caches. Using the proxy cache, we can avoid the bottleneck bandwidth, reduce the transport delay and achieve high quality multimedia services.

The characteristics of real-time video streaming data and the transport requirements are discussed in this paper. A new framework of multimedia proxy caching is presented. Some key technologies as the prefetching management, the replacement management, and the resource scheduling of different media are also studied.

1 The Requirement of Video Streaming Transport over the Internet

The transport of real-time video streaming over the Internet has many characters as follows [3]:

①The size of the video streaming data is very large, and varies with time.

②The video playback is very sensitive to the loss, the error, the delay of the video streaming data on the transmission.

③There exist different importance on the different part of the video streaming data.

④The transmission of video streaming allows error on some degree, even exchanges the efficiency and the quality with some local errors.

⑤The heterogeneous of the network and the clients.

To satisfy the requirement of the transmission of the real-time video streaming, four guarantees must be made. Firstly, the transport channel must provide minimum bandwidth requirement to achieve the acceptable presentation quality. Secondly, to guarantee the continuously playing out of the video streaming, the transmission time must be less than the bounded end-to-end delay. Thirdly, to achieve the desired visual quality, loss of the packets must be kept below a threshold. Finally, because different sub-networks of the Internet have different processing capabilities, different bandwidth, different storage, etc. different clients have different requirements on the quality of the video. How to satisfy these things is a very challenging problem.

The current Internet can only support best-effort services, which does not provide bandwidth reservation for the video streaming, cannot guarantee the timely transport of the video, and cannot guarantee the ratio of the loss. Therefore, the quality of the video cannot be up to the clients' requirement, and the transmission of the video streaming will give burden to the transmission channel and the video server.

The proxy caching technology is introduced to solve these problems. There are many advantages by using this technology.

①Some popular streaming media data can be placed on the proxy cache near the client to reduce the amount of the transmission over the Internet, which will relax the transport compression of the Internet, and lessen the loads of the video server.

②By using the proxy caching technology, we can save the bandwidth of the Internet, and reduce the cost of the transmission.

③The streaming media data can arrive to the clients more quickly and reliably, so the delay of the packet is reduced, and the playback quality of the video is enhanced.

2 The Framework of the Multimedia Proxy Caching

2.1The Disadvantage of the Web Caching Technology on the Video Streaming

The traditional Web Caching handles mostly the non-continuous Web objects, which have small size, such as the html pages, the images. Therefore the granularity to save and to replace is the whole object. But the streaming media object is continuous, and has the very large size. If it is handled like the non-continuous object, there will be only a few streaming media objects in the cache, which will reduce the hit ratio of the cache, and damage the performance of the cache. Therefore, the idea of partial caching is put forward, in which only the important and most popular part is placed in the cache. The efficiency and the hit ratio of the cache can be enhanced with this solution, so that the playback quality of the video will be enhanced and the transmission cost will be reduced.

When the streaming media object is added into the cache, proxy cache handles not only the non-continuous object (say, text, image, etc) but also the continuous object (say, video, audio, etc). It is very difficult to handle different types of the data. Firstly, the data in the cache has different types and different characteristics, for example, the real-time streaming media is sensitive to the delay, but can bear the errors at some degree. The Web objects cannot bear any faults, but can allow some delay. Secondly, streaming media have the continuous and real-time characteristics, which need new functionalities to be added in the cache. Finally, the heterogeneity of the sub-networks and the clients need to be considered in the design of the multimedia proxy cache [4,5]. The current Web caching technology is designed for the non-continuous Web objects, it cannot satisfy these requirement above. Therefore, some enhancements must be made to the architecture and some key technologies.

2.2 The Framework of the Multimedia Proxy Caching

Some enhancement has been made to the architecture of the Web caching. The resource management module and the cached data-scheduling module are added. Based on the characteristics of the streaming media data, the cache misses part, the prefetching part, and the replacement part are modified. Fig.1 shows the framework of the multimedia proxy caching. The architecture of the multimedia proxy caching is made up of two modules, request management module and the resource management module.

The request management module handles all the clients' request, and deals with the continuity of the video streaming transmission. There are five parts in the request management module: prefetching management, cache miss management, bandwidth monitor, request scheduling, and return scheduling. The cache miss part is responsible for the request to the original media server, when the requested object is not in the cache. The prefetching part will prefetch the following segments that do not exist in the cache, while transporting the streaming media segments in the cache. By prefetching these missing segments from the server ahead of time, the delivered quality will be maximized. Bandwidth monitor part estimates the avalible bandwidth between the client and the proxy server, and the avalible bandwidth between the proxy server and the original multimedia server, so as to efficiently distribute the bandwidth among the requests. Request scheduling part distributes the bandwidth among the requests from the proxy server to the original multimedia server and schedules their sending order. Return scheduling part distributes the bandwidth among the requests from the proxy server to the clients and schedules the sending order.

The resource management module receives the video segment from the video server, decides the ratio of the cached data. If the available space in the cache is not enough, it need to decide which segment will be deleted. The resource management module is made up of two parts. The query part receives the clients' request, and looks for the requested data in the cache. If those exist, the requested segment will be sent to the client under the management of the request management module. If not, the cache miss message will be sent to the request management module. The replacement management part decides which segments to be saved in the cache when the useful space is not enough; it also decides which segments will be deleted.

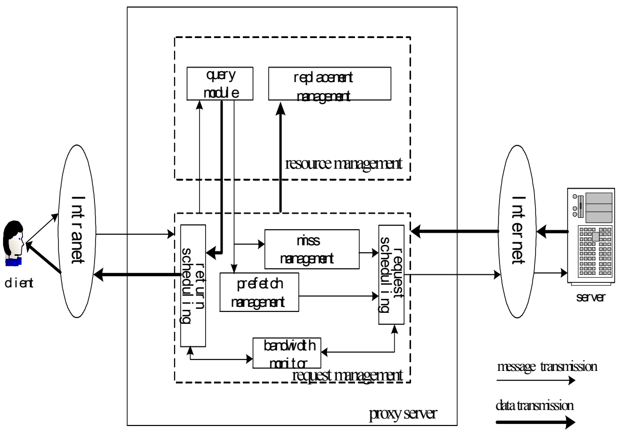

Fig.1 The framework of multimedia proxy cache

3 The Key Technologies of the Multimedia Proxy Caches

3.1 Cache Miss and Prefetching Management

On a cache miss, the request will be relayed to the original server. The proxy server decides whether the video server can transport the data on time by the RTT (Round Trip Time). If not, the proxy server will discard the request. When the request arrives at the origin server, a transmission channel is built up. The video server performs congestion control and quality adaptation based on the state of the session between the source and the client. The client cannot benefit from the cache on a cache miss.

On a cache hit, the proxy server initiates the playback of the request stream instead of the original server. The client will get a shorter startup latency and better playback quality. Since the video streaming is continuous, when handling the request, the proxy cache should prefetch the following missing video segments from the origin server [6] so as to playback the streaming media continuously and smooth out variations in quality. The proxy cache uses the prefetching window to decide the length of the video streaming to prefetch, and uses the useful bandwidth between the proxy server and the video server to decide the transmitted quality.

3.2 Replacement Management

When the video streaming arrives at the proxy server from the original server, some data need to be deleted from the cache using the replacement algorithm if the available space is not enough. The current replacement algorithm [7] treats the data as the whole object, it chooses to reserve or to delete in its entirety based on the popularity of the object. As is mentioned earlier, the multimedia streams can adjust the quality by adding or deleting the layer using the layered video coding. Therefore, in the replacement algorithm, we can only delete the enhancement layer and hold the base layer in the cache to get a higher byte hit ratio. To hide the startup latency, it is preferred to keep the initial part of a layer in the cache.

The client will only request part of the video object. The more the client requests, the more popular the video will be. Let $B_r(i,m)$ be the requested length of layer i of video m and $B(i,m)$ be the length of layer i of video m. During the time t, the popularity of layer i of video m is expressed as follows.

$$p(i,m) = \sum_{x=t-\Delta}^{t} \frac{B_r(i,m)}{B(i,m)} \qquad (1)$$

In the layer-encoded video, there exists the dependence between the layers. The higher layer can only be decoded on the base of all the layers under it. So the lower the layer, the more important the layer is. Lower layers of an unpopular stream might be deleted before the higher layers of the popular stream. The definition of the weighed popularity is introduced, which is that the weighed popularity of layer m is the weighed sum of the popularity of m and the popularity of all the layers higher than m, as follows:

$$p_w(l,m) = \sum_{i=l}^{L} w(i) \times p(i,m) \qquad (2)$$

where w(i) is the weight of layer I, L is the sum of the layer of video m. Using Eq.(2), the popularity of video streams and the importance of the layer can be combined efficiently and the performance of the replacement algorithm is maximized.

The traditional way to test the performance of the replacement algorithm is up to the hit ratio. The higher the hit ratio, the better the performance of the cache replacement algorithm. When the multimedia streams are added in the cache, the aim of the cache is expanded to the reducing of the initial startup delay, the continuous of the playback, and the reducing of the load of the server and the transmission channel. Therefore, the performance of the cache is evaluated not only from the hit ratio, but also from the enhancement of the transmission quality and the reducing of the amount of the transmission, etc.

3.3 The Mixed Media Scheduling

There are many types of applications on the Internet, so the data transported over the Internet include the continuous data and non-continuous data. As is mentioned above, different types of the data have different characters and different requirement to the cache. For example, multimedia streams are very sensitive to the delay but can bear some degree of errors, while the HTML pages need the reliable transmission but can bear the delay. Therefore, how to efficiently cache both continuous and non-continuous data is a challenging problem. The mixed media scheduling module need to specify the corresponding policy of resource distribution and scheduling based on the character of the media. The proxy cache implements management to the mixed media. Under the restriction of the space of the cache and the network bandwidth, it can utilize the bandwidth efficiently and enhance the transmission quality.

4. Conclusions

The character and requirement on the streaming media transmission over the Internet is discussed in this paper. Multimedia proxy caching is presented to alleviate the network congestion and enhance the playback quality. We analyze the deficiency of the Web caching to handle the continuous objects. A new framework is put forward which is suitable to both the continuous objects and the non-continuous objects. We also discuss the key technology in the request management module, resource management module and the mixed media scheduling module. The future work is to realize the system and test the performance of the multimedia proxy cache.

References:

[1] Wu Guoyong, Qiu Xuegang, Wan Yanzai. The video streaming technologies and application over the Internet [M]. Beijing: Beijing University of Posts and Telecommunications Press, 2001.33-34.

[2] Wu Dapeng, Hou Yiwei, Zhang Yaqin. Transporting real-time video over the Internet : challenges and approaches[EB/OL]. http://www.cmlab.csie.ntu.edu.tw/~pkhsiao/PDF/Transporting%20real-time%20video%20over%20the%20Internet-challenges%20and%20approaches.pdf, 2000-08-16/2002-06-12.

[3] Wu Dapeng, Hou Yiwei, Zhu Wenwu, et al. Streaming video over the Internet: approaches and directions[EB/OL]. http://citeseer.nj.nec.com/wu01streaming.html, 2001-12-16/2002-09-08.

[4] Rejaie R, Yu Haobo, Handley M, et al. Multimedia proxy caching mechanism for quality adaptive streaming applications in the Internet[R]. No. 99-709. Dept. of Comp. Sci.,

Univ. of Southern California. 1999.
[5] Rejaie R., Yu Haobo, Handley M, et al.Proxy caching mechanism for multimedia playback streams[A]. San Diego: Internet Proceedings for the Fourth International WWW Caching Workshop. 1999.
[6] Rejaie R, Handley M, Yu Haobo, et al. Proxy caching mechanism for playback streams in the internet[EB/OL]. http://workshop99.ircache.net/Papers/rejaie-html/index.html, 2000-02-09/2002-11-27.
[7] Wang J. A survey of web caching schemes for the Internet[J]. ACM Computer Communication Review,1999,29(5): 36-46.
[8] Williams S, Abrams M, Standridge C, et al. Removal policies in network caches for worldwide web documents [EB/OL]. http://citeseer.nj.nec.com/williams96removal.html, 1999-03-20/2002-05-09.
[9] Ghandeharizadeh S, Zimmermann R, Shi W, et al. Mitra: A scalable continuous media server [J]. Multimedia Tools and Applications Journal, 1997,5(1):79-108.