Title:QoS based retrieval strategy for video on demand

Authors:Panigrahi Niranjan,Sahoo Bibhudatta

Citation:Proceedings of RAIT-09,6-7Feb.2009,ISM University, Dhanbad,Pages:263-269

Abstract:Video on demand (VOD) is an emerging multimedia application that provides a service to users to browse and watch videos as per their wish. One of the most challenging aspects in such system is to serve the user in minimum waiting time which is one of the important, application level quality of service (QoS) parameter. If this has to be mapped with the system level parameter, then the appropriate parameter will be the access time of the video blocks which are distributed in different servers. In this paper, we present a novel retrieval strategy which minimizes the access time of the distributed video blocks and hence, provide a better waiting time to the user.

Keywords:Video on Demand, Quality of Service, waiting time, access time, retrieval strategy

Quality Of Service (Qos) Based Retrieval Strategy For Video On Demand

Niranjan Panigrahi[1] Bibhudatta Sahoo[2],
Department of Computer Science & Engineering
National Institute of Technology Rourkela,
ROURKELA, ORISSA, INDIA, PIN-769 008
E-mail: [1]er.niranjan@gmail.com , [2]bdsahu@nitrkl.ac.in

ABSTRACT

Video on demand (VOD) is an emerging multimedia application that provides a service to users to browse and watch videos as per their wish. One of the most challenging aspects in such system is to serve the user in minimum waiting time which is one of the important, application level quality of service (QoS) parameter. If this has to be mapped with the system level parameter, then the appropriate parameter will be the access time of the video blocks which are distributed in different servers. In this paper, we present a novel retrieval strategy which minimizes the access time of the distributed video blocks and hence, provide a better waiting time to the user.

Keywords: Video on Demand, Quality of Service, waiting time, access time, retrieval strategy

1. INTRODUCTION

Distributed multimedia applications, enabled by rapid advances in high-speed network, storage service and switching technologies, are attracting a great deal of academic and commercial interests. Video-on-demand **is** an emerging distributed multimedia application that has recently received much attention from the entertainment, education, telecommunication, and computer industries. It is a system that can provide service in real-time playback of videos according to users' requests through networks. A good quality video-on-demand system has the following requirements [6]**:** a jitter-free video display, synchronized video and audio, a video playback rate of at least 30 frames per second and complete user control over viewing. The high-bandwidth requirement and real-time delivery constraints of video playback are best addressed by LAN based file servers located close to playback clients, while the cost effectiveness of large tertiary storage devices and the desire to share video widely motivate the use of a central repository. What is needed is a way to preserve the cost effectiveness of centralized storage while maintaining the high performance and scalability of distributed servers [14].

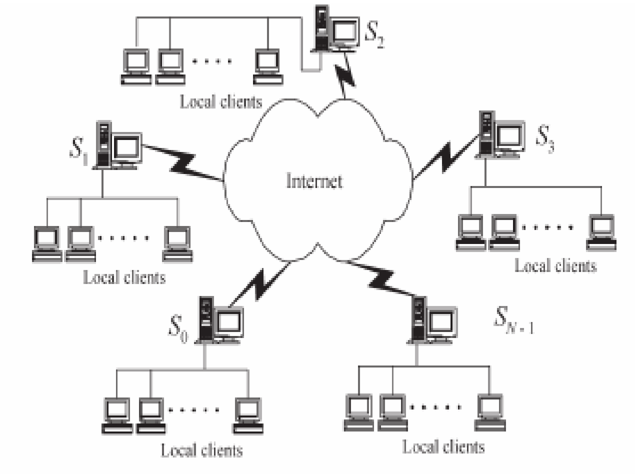
The rest of the paper is organized as follows. Section 2 discusses about the related work in the area of video on demand. Section 3 outlines the distributed architecture for video on demand. Section 4 highlights the theory behind the retrieval process. Section 5 presents the novel retrieval strategy. Section 6 shows the performance evaluation through simulation. Finally section 7 concludes the paper.

2. RELATED WORK

Owing to ever increasing user demands, when the user access rates increase, several issues need to be tackled, e.g., high block rate, long startup delay, glitch (i.e. service interruption), jitter (e.g., frame losing) etc.[2]. In detail, common focus includes, scheduling of multimedia streams for minimizing access times in broadcasting [1], network structure optimization for minimizing the transmission cost, disk scheduling for minimizing access times[8][9][10][11], while maximizing the number of continuous streams and minimizing the buffer space requirement , movie placement strategy in a disk array for minimizing the number of disks [12][13], storage hierarchy optimization for minimizing the storage cost [6], movie buffer caching strategy for effectively utilizing the memory and reducing disk I/O overheads, and dynamic network resource allocation for improving transmission rates with low jitter variation in media streams [7]. Among these researches, the design of efficient strategies to retrieve multimedia documents such as video clips, movies (short and long durations), etc, is a very important research area [3].

3. VOD SYSTEM ARCHITECTURE

To serve each user, the server system retrieves and delivers video at the video playback rate (e.g., 1.5 Mb/s for MPEG video). Consequently, the server system should have a large capacity to handle a large volume of bits. To satisfy this requirement, the server system can adopt one of two configurations, (i) centralized server configuration and (ii) distributed server configuration [4]. Serpanous and Bouloutas [5] made a comprehensive comparison between these two configurations. In particular, the distributed server configuration is attractive in three aspects:(i) Good scalability (ii) High availability and (iii) Competitive performance-to-price ratio For our problem domain, the distributed multi-server video on demand system is considered as shown in Fig.1. Upon the arrival of a request, a server seeks the requested movie locally first. If this movie is available locally, then the movie is retrieved and presented to the user. However, if the requested movie is not available locally, this original server can obtain the information about the requested movie on other servers by employing look-up services, such as the directory service. Then, the requested movie can be retrieved from one or more servers employing our proposed strategy



.Figure 1: A typical video on demand architecture

4. RETRIEVAL PROCESS FOR VIDEO ON DEMAND

Let a request for a movie is placed locally at the proxy server/scheduler and let the requested movie, of size L bits, be present at servers S_0, S_1… S_n. Let the connection bandwidths of channels from other servers be denoted as bw_i, $i = 0, 1, 2…n$, measured in bits per second, and let the playback rate at the client site be R_p, measured in bits per second. From each server, a portion (block) of the entire movie, denoted as m_i, $i = 0, 1, 2…n$ is retrieved and is collected by the proxy in a particular order. Upon receiving the first portion of the movie from S_0, the playback may start at the user terminal, as retrievals from other servers are underway. Thus, in order to start the playback when retrievals from other servers are underway, the size of the portion retrieved must be such that there should not be any data starvation for playback. In other words, the size of the portion retrieved must guarantee the presentation continuity. Now, the retrieval strategy must be such that before the playback of this portion comes to an end, the next portion of the requested movie data should be made available from S_1 and so on.

5. THE RETRIEVAL STRATEGY

The retrieval process is initiated not by retrieving a whole portion (block) that is stored in a server but by retrieving a reasonable portion with the help of which the playback can be started at the client side. This sub-portion is known as critical size of that portion. So, the strategy starts with determination of the critical size.

Step: 1. Determination of critical size

Consider a scenario in which a portion of the movie of size m is to be retrieved from a server using a connection bandwidth bw demanding a playback rate of Rp at the client site. The client can safely start playing the portion after the critical size cs of this portion has been retrieved. In order to guarantee a continuous playback, the time to retrieve the remaining portion (m − cs) must not be greater than the entire playback duration of the portion m. In other words

$$\frac{m}{Rp} \geq \frac{m - cs}{bw} \qquad (1)$$

$$\Rightarrow cs \geq \frac{(Rp - bw)\, m}{Rp} \qquad (2)$$

Step : 2. Precedence relationship between retrieval and playback node

The retrieval process can be represented as a directed flow graph where the arrows capture the precedence relationships in the retrieval and playback portions. For example, portion i can be played after portion ($i - 1$) and after receiving its critical size. From figure 2, we can derive a relationship between the retrieval of portion i and ($i + 1$) and the playback time of the portion mi, with the use of the causal-precedence relation and continuity constraint as

$$\left(\frac{csi}{bwi}\right) + \left(\frac{mi}{Rp}\right) \geq \frac{csi + 1}{bwi + 1} \qquad .... \quad (3)$$

By using the expression for csi from (2) in expression (3), we have

$$\frac{((Rp - bwi)mi)}{Rp\ bwi} + \frac{mi}{Rp} \geq \frac{(Rp - bwi+1)mi+1}{Rpbw\ i+1} \quad \text{....(4)}$$

=>

$$mi+1 \leq \frac{(Rpbwi+1)mi}{(Rp - bwi+1)\,bwi} \quad \text{....(5)}$$

Let us denote $Rpbwi+1/(Rp - bwi+1)\ bwi = \rho i$. Rewriting (5), we have

$$mi+1 \leq mi\rho i \qquad i = 0, 1, \ldots, N-2 \qquad \ldots(6)$$

Step :3. Determination of minimum access time

The use of equality relationships in (5) and (6) results in the maximum size of all the portions other than $m0$. But sum of portions stored at the servers is equal to the length of the movie. i.e.

$$\sum_{k=0}^{N-1} mk = L \quad \text{....(7)}$$

Hence, using (7), we obtain a minimum value for $m0$, equivalently the minimum cs0. In other words, we obtain a minimum access time. Each mi can be expressed in terms of $m0$ as

$$mi = m0 \prod_{k=0}^{i-1} \rho k \ , \quad i = 1, 2, \ldots, N-1. \qquad \text{....(8)}$$

Thus, the above set of $(N - 1)$ equations given by (8), together with (7), are solved to obtain the individual disjoint portions of the requested movie. Substituting each mi from (8) into (7), we obtain

$$m0 = \frac{L}{1+\sum_{p=1}^{N-1} \prod_{k=0}^{p-1} \rho k} \qquad \text{....(9)}$$

Substituting (9) in (8), we obtain the individual sizes of the portions as

$$mi = \frac{L \prod_{k=0}^{i-1} \rho k}{1+\sum_{p=1}^{N-1} \prod_{k=0}^{p-1} \rho k}, i = 1, 2, \ldots, N-1 \qquad \text{....(10)}$$

Thus, the minimum access time is given by

$$ATmin = \frac{cs0}{bw0} = \frac{(Rp - bw0)m0}{Rpbw0} = \frac{L\left(\frac{1}{bw0} - \frac{1}{Rp}\right)}{\left(1+\sum_{p=1}^{N-1} \prod_{k=0}^{p-1} \rho k\right)} \qquad \text{....(11)}$$

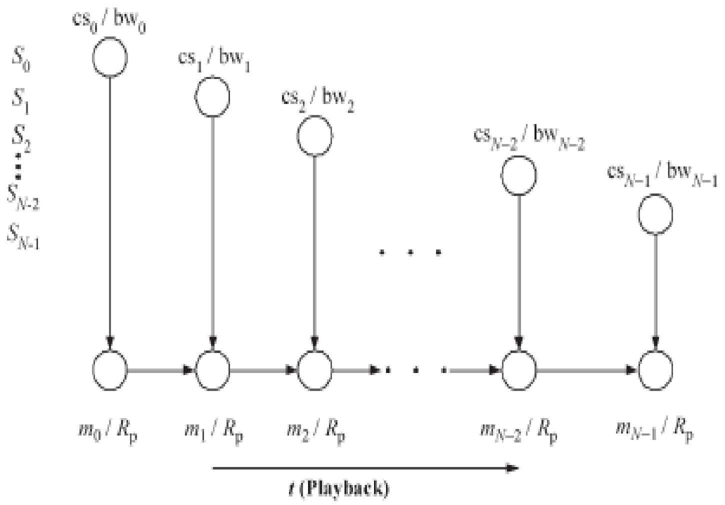


Figure 2: Directed flow graph representing the retrieval process

6. PERFORMANCE EVALUATION

The performance of the retrieval strategy is evaluated for the system that handles one client request, identical connection bandwidths, i.e., bw_i = bw=1 Mbps for all the channels. The movie size *L* is assumed to be 2 Gb, and the playback rate *R*p is 1.5 Mb/s (MPEG-I standard)..From fig 3, it is observed that the given strategy remarkably outperforms the PAR (Play After Retrieval) strategy [1] in minimizing the access time.

No. of servers	Access time in Seconds (PAR strategy)	Access time in Seconds (Proposed strategy)
2	585.14	170.66
3	210.05	52.51
4	80.70	17.06
5	31.78	5.64
6	12.63	1.87
7	5.04	0.62
8	2.01	0.20

Table 1:Acess time of PAR and Proposed strategy

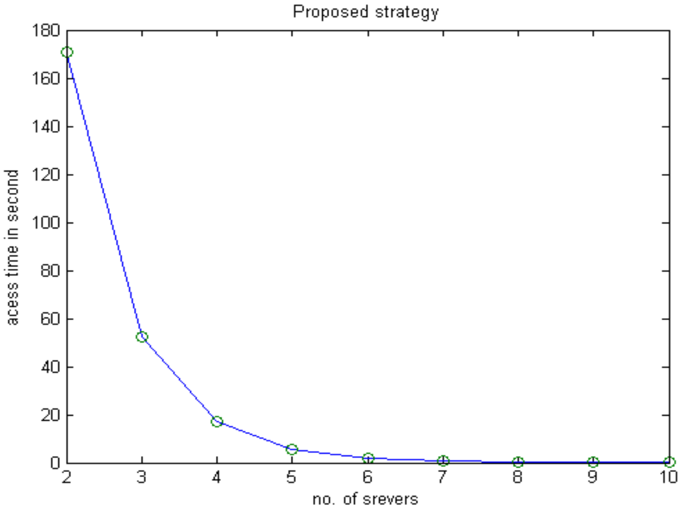


Fig 3(a): Proposed strategy

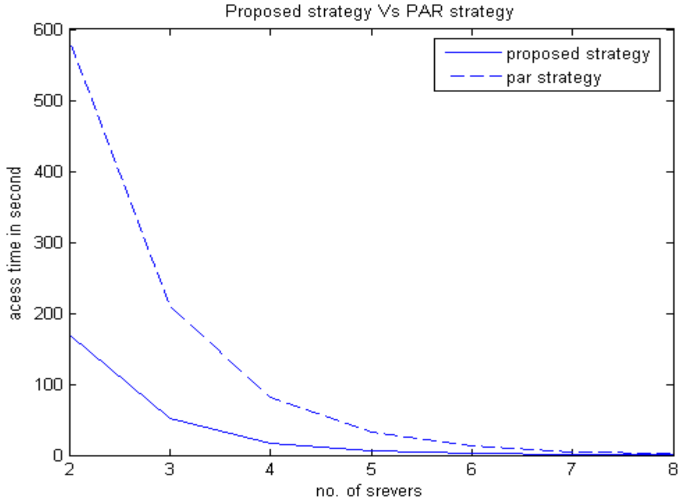


Fig 3(b): Proposed strategy Vs PAR strategy

7. CONCLUSION

The strategy presented in this paper is only based upon the retrieval theory under homogeneous environment; however there are important related issues that can be considered as natural extensions to the current design. These include retrieval under heterogeneous environment (e.g. different connection bandwidth between servers), multiple client requests under varying traffic conditions, coordination of servers during retrieval, admission-control mechanism, implementation of VCR-like operations, and combined use of strategies such as batching, patching, and caching, with the strategy for better resource management.

REFERENCES

[1] B. Veeravalli and G. D. Barlas, “Access time minimization for distributed multimedia applications,” Multimedia Tools Appl., vol. 12, no. 2/3, pp. 235–256, Nov. 2000.
[2] D. Ligang, V. Bharadwaj, and C. C. Ko, “Efficient movie retrieval strategies for movie-on-demand multimedia services on distributed networks,” Multimedia Tools Appl., vol. 20, no. 2, pp. 99–133, 2003.
[3] Long Chen and Bharadwaj Veeravalli,” Multiple-Server Movie-Retrieval Strategies for Distributed Multimedia Applications:A Play-While-Retrieve Approach”, IEEE trans. on systems, man and cybernetics—part A: systems and humans, vol. 36, no. 4, pp. 786-803, July 2006
[4] Yiu-Wing Leung and Ricky Yuen-Tan Hou, ” Assignment of Movies to HeterogeneousVideo Servers” IEEE trans.on systems,man, and cybernetics—part A : systems and humans,vol. 35,no. 5, pp. 665-681, Sept 2005
[5] D. N. Serpanos and A. Bouloutas, “Centralized versus distributed multimedia servers,” IEEE Trans. Circuits Syst. Video Technol., vol. 10, no. 8, pp. 1438–1449, Dec. 2000.
[6] Chor Ping Low, Hongtao Yu, Jim Mee Ng, Qingping Lin, Yacine Atif,” An Efficien Algorithm for the Video Server Selection Problem”, Global Telecommunications Conference, 2000. GLOBECOM '00 ,vol 3,pp:1329-1333
[7] Hongtao Yu, Chor Ping Low, Yacine Atif,” Design Issues on Video-on-Demand Resource Management”,Proceedings of IEEE International Conference on Networks (ICON 2000). 5-8 Sept. 2000 Page(s):199 - 203
[8] Putra Sumari, Azman Samsudin, Hailiza KamaruIhaili,” Data Storage and Retrieval for Video-on-demand Servers”, Proceedings of the IEEE Fourth International Symposium on Multimedia Software Engineering (MSE’02)
[9] Robert Flynn, William Tetzlaff,” Disk Striping and Block Replication Algorithms for Video File Servers”, international Conference on Multimedia Computing and Systems (ICMCS '96), 1996, IEEE Computer Society, Washington, DC, 0590.
[10] J. Aerts, J. Korst, and S. Egner, “Random Duplicate Storage Strategies for Load Balancing in Multimedia Servers,” Information Processing Letters, vol. 76, nos. 1-2, pp. 51-59, 2000.
[11] Tat Seng Chua ,Cheng Hian Goh, Beng Chin Ooi , Kian-Lee Tan,” A Replication Strategy for ReducingWait Time in Video-On-Demand Systems ”, Multimedia Tools and Applications, 15, pp.39–58, 2001
[12] James Z. Wang, Ratan K. Guha,” Data Allocation Algorithms for Distributed Video Servers”, In Proceedings of the Eighth ACM international Conference on Multimedia. MULTIMEDIA '00. ACM, New York pp:456-458.
[13] Chor Ping Low,” An efficient retrieval selection algorithm for video servers with random duplicated assignment storage technique”, Information Processing Letters, vol.83, pp. 315-321, 2002,www.elsevier.com/locate/ipl
[14]A. Narasimhan, "Distributed multimedia applications-opportunities, issues, risk and challenges: a closer look", pp.455-460, 1997 IASTED International Conference on Intelligent Information Systems (IIS '97), 1997