Techniques for Improving the Capacity of Video-on-Demand Systems

Hari Kalva
Center for Telecommunications Research
Columbia University,
New York, NY 10027

Borko Furht
Multimedia Laboratory
Florida Atlantic University
Boca Raton, FL 33431

Abstract

In this paper we present two techniques to improve the capacity of video on demand systems. Video on demand is an electronic video rental system in which clients request and play videos on-demand. Video on demand system can be implemented over an existing cable TV network or an upgraded ADSL network. The two techniques used to improve the capacity of video on-demand systems are segmentation and multicasting. Segmentation consists of dividing the video into several fixed length segments, and then transmitting the segments at regular intervals instead of transmitting the entire video continuously. Multicasting assumes that each subscriber has a limited storage space, so same video segments can be multicast to subscribers simultaneously. Results of a comprehensive simulation study presented in the paper show a significant improvement in the capacity of the system when these two techniques are applied. In evaluating video on-demand systems we considered the following parameters: the number of users supported, the number of videos played per day, and the number of requests rejected per day.

1. Introduction

Advances in computer and communication technologies have resulted in a number of new multimedia services. Important applications benefitting from these technological advances are interactive services such as video conferencing, video on-demand, home shopping networks, distance learning, and collaborative computing. Of these services, video on-demand (VOD) is the one with immense commercial potential and will be one of the first multimedia services to enter the home environment. Sophisticated transmission techniques enable high speed data transmission over existing telephone and cable TV networks and compression techniques such as MPEG-1 allow the transmission of motion video at a rate of 1.5 Mbps. With today's cable TV a user has little or no control over the programs he/she watches. With VOD, users will have the flexibility of selecting the content as well as scheduling the program they want to watch.

The rest of the paper is organized as follows. Communication architectures for VOD services are discussed in section 2. The segmentation and multicasting techniques are presented in sections 3 and 4 respectively. The simulation model is described in section 5, simulation results are presented in section 6, and conclusions are presented in section 7.

2. Communication Architectures for VOD Services

A typical VOD system consists of a video server with digitally stored videos stored on high capacity storage devices such as optical disks and a communication network connecting the users to the server. The components of a typical VOD system are shown in Figure 1. The Customer Premises Equipment (CPE) consists of a set-top box and a monitor. The functions of a set-top box include decoding a compressed video, demodulation, de-scrambling, program storage, and others [1].

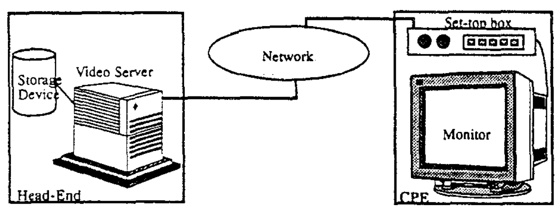

Figure 1. Components of a VOD system

A distributed VOD system consists of a hierarchy of multimedia archives connected by high speed wide area networks (WANs) and metropolitan area networks (MANs) as shown in Figure 2. Several architectures have been proposed for VOD services [2, 3, 4, 5]. Various issues involved in the design of VOD systems are discussed in [6]. Bell Atlantic's VOD architecture discussed in [4] uses a Asymmetric Digital Subscriber Loop (ADSL) network at 1.536 Mbps to transmit the digitized video. ADSL-II can support data rates of up to 6 Mbps. In addition to an ADSL network, the network that has potential to provide VOD services to home is Community Access TV (CATV) network. Capacity analysis of a CATV network for multimedia on-demand is presented in [7]. In most of the VOD architectures discussed in the literature, the videos are transmitted continuously at a rate of 1.5 Mbps, assuming that videos were compressed using MPEG-1. This would tie up the channel for the entire duration of the video play back and would limit resource sharing. A delivery scheme to multicast a video is presented in [8]

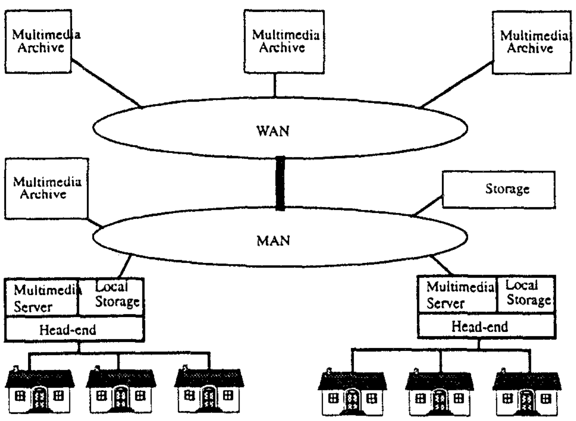

Figure 2. Distributed multimedia on Demand System

The network that is most likely to be used to deliver VOD is CATV network. CATV network uses a coaxial cable and has very large bandwidth that can be used for VOD services. The coaxial cable used in CATV networks has a bandwidth of about 750 MHz. Systems that allow 10 Mbps ethernet connections over a 6 MHz channel are commercially available. Advanced modulation techniques such as 64 QAM and 16 VSP will allow data rates of 30 and 43 Mbps respectively, on a 6 MHz channel. This allows a maximum bandwidth of 4.7 Gbps from head-end to the user. Most of this bandwidth will be under utilized if a video is played continuously. The capacity analysis of a VOD system using a CATV network is presented in [7]. The model used in [7] consists of a CATV tree with N users connected to the head-end. The users request videos from the server at the head-end. The request rate as seen by the head-end is modeled using normal distribution and is given by the following equation:

$$(t) = \frac{N \bullet \frac{\lambda}{7}}{\sigma\sqrt{2\pi}} e^{-(t-\mu)^2/(2\sigma^2)} \qquad \text{(eq. 1)}$$

The parameters used in equation (1) are:

$\lambda_G(t)$ - request rate as seen by the head-end at time t.

N - number of users connected to the head-end.

λ - number of Videos requested per User per Week (V/U/W).

σ - mean length of the videos.

μ - peak hour.

The requests are served on a first-come-first-served basis and are not queued. When a request is accepted, a channel is reserved for the connection for the entire duration of the video. The number of channels available for VOD is assumed to be 450. The access network between the head-end and CPE is assumed to be a CATV network. In additions to these assumptions we further assumed that:

- User has a limited buffer so that video can be pre-delivered.
- A control channel between user and head-end is available during video play back.
- A video is segmented in to S fixed length segments of length T_s minutes.
- Videos can be transferred at a rate 3 times the play back rate.
- All videos available are equally popular.

In a VOD system the main goal is to support the maximum number of users possible with minimal waiting time, as waiting can be often annoying to users. If a channel is not available when a request is made, the request is rejected. Since there is no queueing of requests in the system considered, all the accepted requests have zero waiting time. Hence the maximum number of users that can be supported by the system being modeled with no rejections will give the maximum number of users supported such that no user will have to wait for a connection.

When a video is transmitted continuously, a channel has to be dedicated for the entire duration of the play back. The main disadvantage of such a connection is that a maximum of C users can be served at any time, where C is the number of channels available for VOD. This drastically limits the number of users that can be supported such that there are no rejections. Table 1 gives the number of users supported by the system with no rejection for different request rates (l) for 60 minute (s = 60) videos. We used the request distribution from [7] (equation (1)) for the purpose of simulation. The simulation is done with a granularity of 1minute, i.e. observations are made after every minute.

Table 1: Maximum number of users supported with no rejection in a simple VOD system (no segmentation, no multicasting)

Requests in videos per user per week (V/U/W) (λ)	Video length in minutes (σ)	Max. users supported (N)
1	60	8150
2	60	4070
3	60	2710
4	60	2040
5	60	1630
6	60	1360
7	60	1170

From Table 1, it is clear that as the number of videos requested per week increases, the number of users that can be supported with no rejection decreases. Thus the number of users supported by the head-end decreases with increasing request rate resulting in an increase in the charges incurred by the user. For a VOD system to be feasible, the number of users connected to a head-end should be large so that user expenses are low.

3. Segmentation Technique

The disadvantage of tying up network resources associated with continuous transmission of a video can be overcome by using the segmentation technique. This technique makes use of the high network data rates and a limited buffer available at customer's premises. We assumed a delivery rate thrice that of the playback rate; i.e. a 6 minute video segment can be transmitted in 2 minutes. The basic transaction units in this scheme are segments; i.e. the videos are transmitted as multiple segments and if there is a need for retransmission, the entire segment has to be retransmitted. First we will analyze the storage requirements for buffering. Using MPEG-1 a video can be compressed to 1.5 Mbps and gives a VHS quality picture. Assuming MPEG-1 compression, a 6 minute segment would thus require a buffer size of 67.5 MBytes. A 202.5 MBytes of storage is required to buffer three 6 minute segments of a video, and this is small enough for practical implementation.

The segmentation technique is illustrated in Figure 3, which shows the transmission of a 10 segment video to a user. The segments are transmitted at regular intervals of 6 minutes. The control connection that exists between the user and the server at the head-end is used to request videos and report the status of the video being played. This technique, when used to transmit a 60 minute video, will tie up a channel for only 20 minutes. The same channel can be used to serve other requests without disturbing the current session. During the playback, it is important to ensure the continuity of playback. If a dedicated channel is used to transmit a video, there will not be any discontinuity during playback. Segmenting a movie will add additional burden of maintaining continuity. In our system, continuity is ensured by reserving the channels for video delivery.

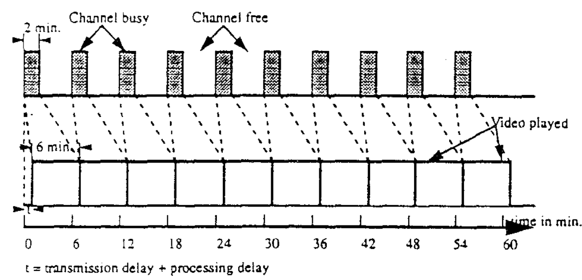

Figure 3. Segmentation technique: a 60 min video is divided into ten 6 min. segments; each segment is transmitted in 2 min.

4. Multicasting Technique

Communication is generally either unicast or broadcast. In addition to these two modes of communication, there is another mode of communication used in group communications called multicasting. Multicasting is a communication technique in which data is sent to a group of users at the same time over a single channel. Multicasting in a VOD system and scheduling algorithms for multicasting are discussed in [8, 9]. In a VOD system there is a possibility that more than one user requests the same video at the same time. The probability of more than one user requesting the same video will be high if the number of videos available is small when compared to the number of users. Even if there is a large video archive, there will be a set of popular videos which are requested by many users, thus increasing the chance of multicasting. If a requested video is multicast, all the users in the multicast group will be served by one channel thus saving the network bandwidth.

In the simplest case of multicasting, a video is multicast to users who request the video at the same time, called synchronized multicasting. When a video is segmented and users have the ability to buffer few segments of a video, multicasting is possible even for asynchronous requests for same video [8]. In such cases only a part of the video is multicast but will result in significant increase in system capacity. The suitability of a request for partial multicasting depends on the number of video segments that can be buffered at users premises. Not all asynchronous requests can be multicast, because, only a limited number of segments can be buffered by a user. Multicasting two sessions with starting times widely apart may result in overwriting the segments that were multicast earlier, causing unnecessary retransmissions. Different multicasting scenarios are shown in Figure 4.

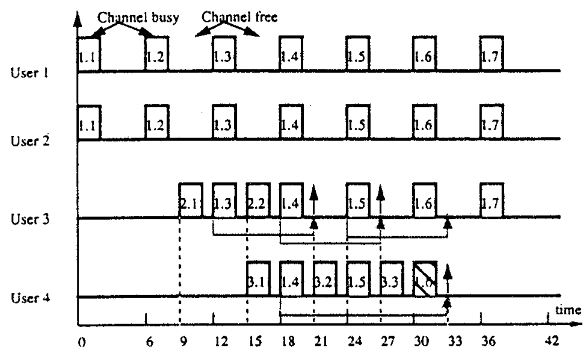

Figure 4. Figure 4. Multicasting scenarios

Users 1 and 2 request a video at the same time (t = 0), and the movie is multicast to them. Both users 1 and 2 are served by a single channel and segments are delivered on channel 1. User 3 requests for the same video at time t = 9 and is scheduled for partial multicasting. The third segment of the video is multicast to the three users and the first two segments of the video are transmitted to user 3 on channel 2 (2.1 and 2.2). The segments pre-delivered to user 3 are scheduled for playback as follows: segment 1.3 at time 21, segment 1.4 at time 27, etc.

The user 4 requests the same video at time t = 15. If this user is scheduled for multicasting, segment 4 of the video multicast on channel 1 will also be received by user 4. User 4 has to receive the first three segments of the video on a separate channel (channel 3). Since segment 4 is not scheduled for playback until t=33 and since the user can buffer only 3 segments, the segment 6 multicast on channel 1 will overwrite segment 4. Since segment 4 cannot be retransmitted before t=33, this will result in unnecessary delays and retransmissions. Hence user 4 is not suitable for multicasting and a fresh session has to be started on channel 3.

If S_n is the number of segments that can be buffered at users' premises, for partial multicasting to be possible, the video has to be requested before segment number S_n is transmitted in the current session. In other words, partial multicasting is possible only if segment S_n is multicast. When a request for partial multicasting is accepted, the first few segments are transmitted on separate channels or multicast among the synchronous requests within. If the number of segments S_n that can be buffered at user's premises is increased, the probability of multicasting is also improved. For example, consider the case depicted in Figure 4, if S_n is increased to 4, user 4 can also be included in the partial multicasting session without overwriting the earlier pre-delivered segments.

5. Simulation Model

The VOD system modeled consists of:

- A head-end with a server and V digitally stored videos. All the requested videos are available at the head-end and transmitted to the user as multiple segments of fixed length T_s.
- A CATV network with 450 channels available for VOD, each with a capacity of 4.5 Mbps.
- CPE consisting of a monitor and a set top box which can buffer up to S_n segments of a video.

Our objective is to determine the number of users supported by a head-end such that all requests are accepted with zero waiting time. The admission policy is simple; if a channel is available at the time of request, request is accepted otherwise it is rejected. We limit the number of users served by a head-end so that no request is rejected.

When a request is accepted at time t, the first segment of the video is transmitted. The second segment has to be transmitted at $t+T_s$, where T_s is the length of a video segment in minutes. To ensure continuity of playback, channels are reserved for the transmission of rest of the video segments. If a session is started for a S_v segment video at time t, a channel will be reserved at times $t+i{\cdot}T_s$, i = $1,...,S_v$. Since the video is transmitted as segments, a channel reserved at time t will be busy for $T_s/3$ minutes (assuming a delivery rate thrice the playback rate) and will be released at time $t+T_s/3$. Since a channel is not tied up for the entire duration of the video, the channel utilization will improve as the number of users served by the same channel is increased.

The motivation behind multicasting a video is the possibility of more than one user requesting a same video. To simulate the effect of multicasting, we assumed that there are V videos in the video storage. When a request is made, a video is randomly assigned to the request. The accepted request can be scheduled for multicast, partial multicast, or unicast. A video is unicast when there is only one request for the video and there is no earlier session of the video that is suitable for partial multicast. A video is multicast if there is more than one request for the same video and there is no earlier session of the video that is suitable for partial multicast. Partial multicasting is possible if there are one or more requests for a video and a session of the requested video was started early enough for current requests to be scheduled for partial multicasting. The first few segments of the requested video are multicast if the number of current requests is more than 1.

If R requests are addressed to a video i, and a new multicasting session is scheduled, only one channel is reserved to serve the R requests instead of R channels reserved in the case with out multicasting. If the R requests are scheduled for partial multicasting with a session that is playing the segment S_p of the requested video, only one channel is reserved to multicast the first S_p segments of the video.

6. Simulation Results

In this section, we present the results obtained using the simulation model described in section 5. Simulation is performed for three VOD systems: (a) a simple VOD system without segmentation and multicasting, (b) VOD system which includes segmentation only, and (c) VOD system which includes both segmentation and multicasting. In all experiments we assumed that the length of the video is 60 minutes. The various parameters which are evaluated and calculated are, the total number of users supported in a VOD system, the average number of videos requested per week by each user (V/U/W), and the number of videos played and rejected per day.

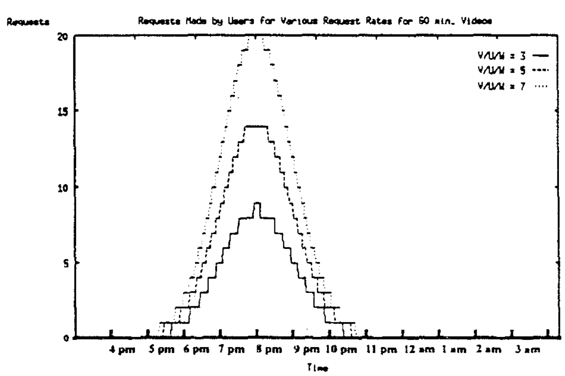

Figure 5. Requests submitted to the head-end

The request rate of the users are modelled by the equation (1), as shown in Figure 5. We assume that the highest request rate is at 8 pm. Three different user request rates are shown in Figure 5 for the number of videos requested by each user per week (V/U/W) equal to 3,5, and 7.

6.1 Simple VOD system (no segmentation and no multicasting)

This system has also been evaluated in [7]. The simulation results obtained, shown in Figures 6 and 7, are similar to those reported in [7]. In Figure 6 we analyzed the performance of a VOD system which supports 3000 users with 60 minute videos. It can be concluded that this system can support 3000 users with an average request rate of only 2.5 V/U/W with no rejection. If the average request rate increases (higher than 2.5), total number of rejections per day increases linearly as shown in Figure 6.

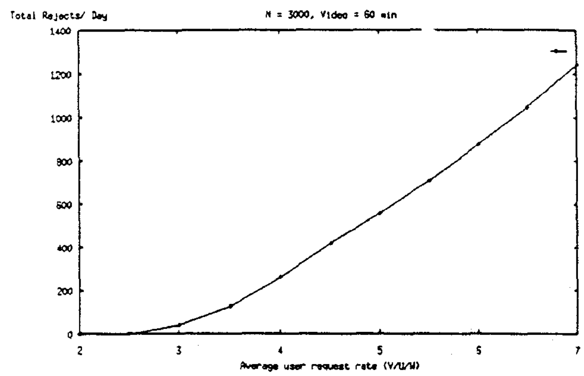

Figure 6. Average user request rate (V/U/W) supported for N = 3000, and 60 min. videos

In Figure 7, we reported the results of simulation where the average number of user requests (V/U/W) was set to be 6 videos per week. We analyzed the total number of users that can be supported by such VOD system. The results from Figure 7 indicate that in this case only 1350 users can be supported without rejecting any request. When the number of users increases over 1350, the total number of rejections will also increase linearly, as shown in Figure 7.

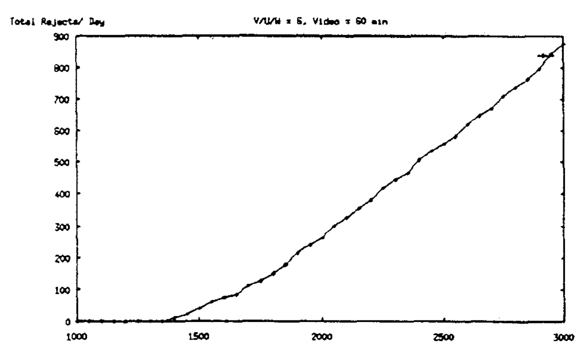

Figure 7. Total number of rejections per day as a function of number of users supported

6.2 VOD system with segmentation

Segmenting a video for transmission will increase the capacity of the VOD system as free channels are available to satisfy the requests. Figure 8shows the request rate in V/U/W supported by the head-end serving 3000 and 5000 users. If the number of users connected to the head-end is small, higher request rates can be supported. A head-end serving 3000 users can satisfy a user request rate of 8.5 V/U/W without rejecting any request. Supporting 5000 users reduces the request rate to 5 V/U/W as shown in Figure 8. Figure 9 shows the maximum number of users supported by the head-end transmitting 60 min. videos at a rate of 6 V/U/W. With segmentation technique, the head-end can now support 4300 users for a request rate of 6 V/U/W and 60 min videos This is an improvement of more than 218% over a similar case without segmentation technique (1350 users; Figure 7).

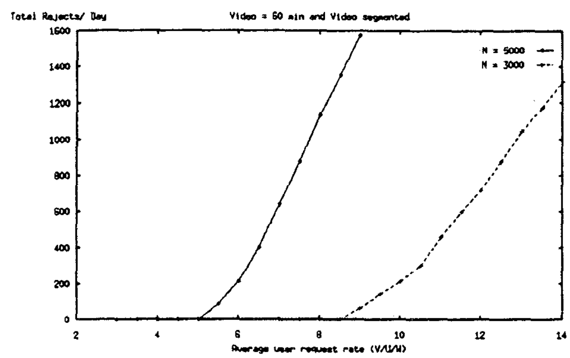

Figure 8. VOD system with segmentation: average user request rate supported

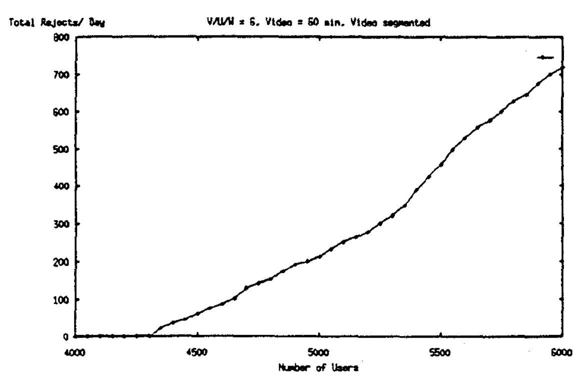

Figure 9. VOD system with segmentation: total number of rejections per day as a function of number of users supported

6.3 VOD system with segmentation and multicasting

In this simulation, we evaluated the effects of both segmentation and multicasting, on the performance of the VOD system. Figure 10 shows the effect of multicasting on the average user request rate supported by a head-end serving 3000 users with 60 min. videos. With a video database of 500 videos, as many as 19 V/U/W can be supported with no rejections. If the number of videos in the video database is increased to 1000, the number of requests supported is reduced to 12 V/U/W. This is due to the decrease in the probability of more than one user requesting the same video as all the videos are assumed to be equally popular.

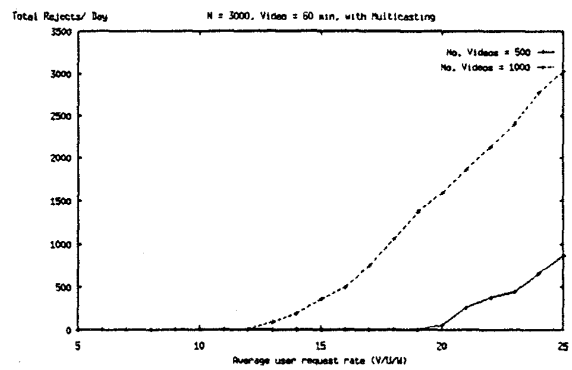

Figure 10. VOD system with multicasting: average user request rate supported for 3000 users

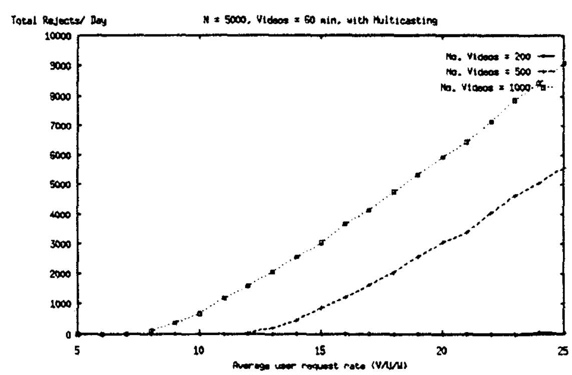

Figure 11. VOD system with multicasting: average user request rate supported for 5000 users

Figure 11 shows the same results for 5000 users. If the number of videos in the database is as low as 200, the probability of more than one user requesting the same video is high and as many as 25 V/U/W can be satisfied by the head-end without rejecting any request.

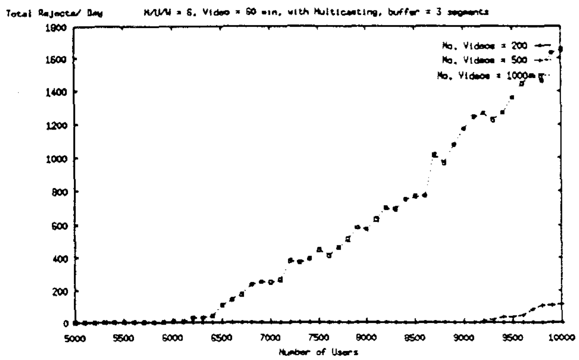

Figure 12. VOD system with multicasting and buffer capacity of 3 segments: total number of rejections per day as a function of number of users supported

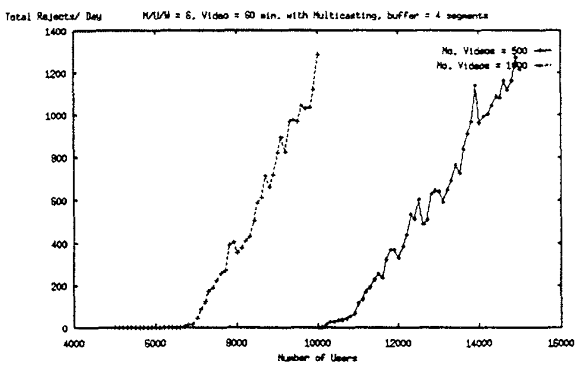

Figure 13. VOD system with multicasting and buffer capacity of 4 segments: total number of rejections per day as a function of number of users supported

Figure 12 shows the maximum number of users supported by a head-end satisfying a request rate of 6 V/U/W with a mean video length of 60 minutes. With 500 videos in the database, as many as 9100 users can be supported, when the number of videos in the database is increased to 1000, the probability of multicasting decreases and the head-end can support about 6000 users. This is a significant improvement over the similar cases without multicasting. If the buffer size of the user is increased to buffer 4 segments, the number of sessions that can be scheduled for partial multicasting increases and this in turn will increase the capacity of the VOD system. The maximum number of users supported by a head-end satisfying a request rate of 6 V/U/W with a mean video length of 60 minutes and 4 segment buffer is shown in Figure 13. With 500 videos in the database, the number of users supported with no rejection increased by 1000 to 10100 and with 1000 videos in the database, the number of users supported increased to 6600

7. Conclusion

In this paper we described segmentation and multicasting techniques for a VOD system and evaluated their impact on the performance of the system. The results obtained are summarized in Figures 14 and 15.

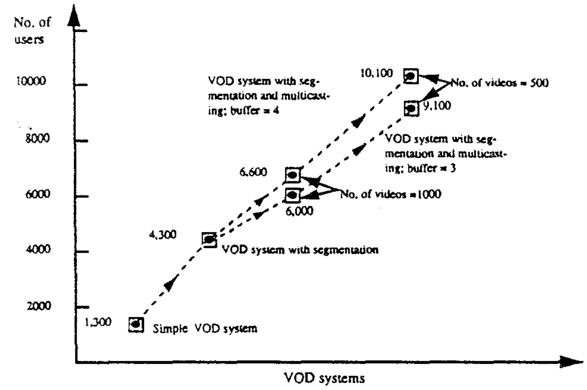

Figure 14. Evaluation of three VOD systems; Criteria: Maximum number of users that can be supported without rejection; average user request rate = 6 V/U/W; video length = 60 min.

Figure 14 shows the increase in the total number of users supported by a VOD system beginning from a simple VOD system (1350 users), to a VOD system with segmentation (4300 users), and finally to a VOD system including both multicasting and segmentation (6000 (6600) users for a database of 1000 videos and 9100 (10100) for a database of 500 videos with a 3 (4) segment buffer at user's premises). In this case, average user request rate is 6 videos per week and each video is 60 minutes.

Figure 15 shows how the average user request rate supported can be increased with segmentation and multicasting techniques, from 2.5 videos per week (simple VOD system), to 8.5 videos per week (VOD system with segmentation), and to 12 and 19 videos per week (VOD system with both segmentation and multicasting for videobase of 1000 and 500 videos, respectively). In this case, the total number of users supported by the VOD system has been 3000 and video length has been 60 minutes.

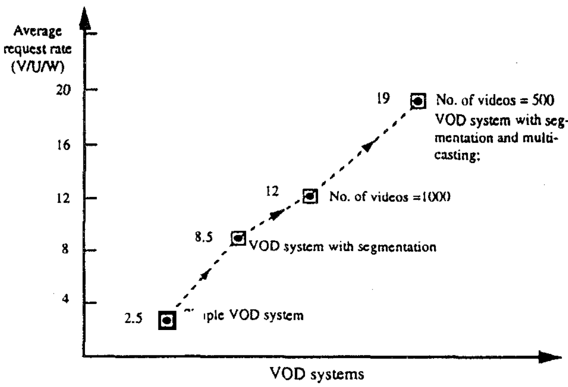

Figure 15. Evaluation of three VOD systems; Criteria: maximum user request rate that can be supported without rejections; number of users = 3000; video length = 60 min

When simulating the multicasting technique we assumed that all videos are equally popular. This is not usually the case; there will be a set of videos in a database which are more popular than the other. This will increase the chances of multicasting resulting in a further improvement in the capacity of the system.

The techniques presented in the paper have an impact in designing a head-end which can support larger number of users, thus reducing the cost per user. In this study, we assumed that all videos are available at the head-end. The future challenging research problem would be to consider the general case in which a video not available at the head-end must be requested and transmitted from the nearest storage server where it is available.

8. References

[1] B. Furht, D. Kalra, and F. Kitson, "Issues in Designing Interactive Television Systems," *Technical Report No. TR-CE-94-29,* Florida Atlantic University, Boca Raton, FL, 1994.

[2] W.D. Sincoskie, "Video On Demand: Is It Feasible," *Proceedings of the 1990 IEEE GLOBECOM*, San Diego, CA, Dec. 2-5, 1990, pp. 201-205.

[3] P.V. Rangan, H.M. Vin, and S. Ramanathan, "Designing an On-Demand Multimedia Service," *IEEE Communications Magazine*, Vol. 30, No. 7, July 1992, pp. 56-64.

[4] J. Sutherland and L. Litteral, "Residential Video Services," *IEEE Communications Magazine*, Vol. 30, No. 7, July 1992, pp. 36-41.

[5] A.D. Gelman *et. al*, "A Store-And-Forward Architecture For Video-On-Demand Service," *Proceedings of the 1991 IEEE International Conference on Communications*, Denver, CO, June 23-26, 1991, pp. 842-846.

[6] T.D.C. Little and D. Venkatesh, "Prospects for Interactive Video-on-Demand," *IEEE Multimedia*, Vol. 1, No. 3, Fall 1994, pp.14-24.

[7] J.P. Nussbaumer, B.V. Patel, and F. Schaffa, "Multimedia Delivery on Demand: Capacity Analysis and Implications," *To be published in the Proceedings of the 19th Local Computer Networks Conference*, Minneapolis, 1994.

[8] H. Woo, J. Park, and C. Kim, "A New VOD Service Scheme with Dynamic Multicasting Capabilities," *Proceedings of the First ISMM Conference on Distributed Multimedia Systems and Applications*, Honolulu, Hawaii, August 1994, pp. 37-40.

[9] A. Dan, D. Sitaram, and P. Shahabuddin, "Scheduling Policies for an On-Demand Video Server with Batching." *Proceedings of the ACM Multimedia 94*, San Francisco, CA, Oct. 15-20, 1994, pp. 15-23.