The Role of Multicast Communication in the Provision of Scalable and Interactive Video-On-Demand Service

Kevin C. Almeroth and *Mostafa H. Ammar*

College of Computing, Georgia Institute of Technology, Atlanta, Georgia 30332-0280
{kevin,ammar}@cc.gatech.edu

Abstract. Multicast delivery can improve the scalability of Video-On-Demand (VOD) systems by allowing multiple customers to share one set of video server and network resources. In such a system, providing customer interactivity can be difficult. In this paper we overview our work which addresses how interactivity may be accomplished in a multicast delivery environment and analyzes the performance of such systems.

1 Introduction

Video-On-Demand (VOD) systems which provide guaranteed interactivity typically allocate one set of video server and network resources for each customer movie request. Such systems may not be scalable because resource usage is directly proportional to the number of customers being served. One alternative is to use multicast communication to share resources and provide service to multiple customers simultaneously. A multicast VOD system servicing multiple customers cannot provide the same individualized, interactive service as a unicast system, but the illusion of one-to-one servicing can still be provided. Typically, when multicast communication is used, some of the "on-demand" nature of VOD may need to be sacrificed[5]; the level of interactivity may need to be reduced; and/or the complexity of system operation and hardware may need to be increased. This paper overviews techniques we considered for using multicast in VOD systems, and summarizes the insights we gained from analyzing such systems.

2 Unicast Video-On-Demand

A unicast VOD system operates by allocating resources and providing interactivity to customers on an individual basis. A customer makes a movie request which is sent to a video server. If a set of video server and network resources are available, the request is satisfied, and playout begins immediately. This type of VOD system services customers in a one-to-one or *unicast* fashion. Resources are reserved on an individual basis, and reservations are typically made independent of other customer requests. Interactivity is provided through VCR functions which can be used to pause, rewind, or fast forward the movie. When a customer initiates a VCR action, a request is sent to the video server where the appropriate change in the playout is made. The advantage of a unicast VOD system is the ability to guarantee the availability of VCR functions. However, because there is a one-to-one relationship between resources and customer requests, the system does not scale well. Furthermore, providing VCR functions may require additional bandwidth[4] and video resources.

3 Multicast Video-On-Demand

One way to more efficiently use resources is to deliver the same movie to multiple customers at the same time. Some of our earlier work examined some of the issues in using multicast communication as an information delivery mechanism[6, 7]. While multicast communication offers the advantage of being scalable, the interactive nature of VOD logically demands individualized service. Several solutions have been suggested which group requests for the same movie. We have investigated solutions based on the use of fixed-length time slots to satisfy multiple requests simultaneously[1, 2]. Others have independently investigated the use of batching and temporal proximity of requests to service multiple requests[3]. In either case, multiple customer requests are satisfied using one set of video server and network resources. Providing interactivity is more difficult in these systems because the video stream multicast to the group cannot be modified because of the VCR actions of one member. Unless tradeoffs in the level of interactivity or system complexity are made, a multicast VOD system must handle VCR actions individually thereby reducing the advantages of using multicast.

3.1 Reducing the On-Demand Nature of the System

In part of our work on VOD systems using fixed-length time slots[1, 2], we examine the tradeoff between the on-demand nature of a system and its scalability. Continuous time is divided into equal length time slots in which customer requests will be grouped. Requests that arrive in the same time slot are serviced together at the end of the time slot. The video server will group the requests, and begin playout of the movie at the end of the time slot. It has been observed that among the set of available movies, a small percentage will be requested very often. For such popular movies, and during prime time, it is likely that many requests for the same movie can be accumulated in each slot. Assuming resources are available, customers need only wait until the end of the time slot in which they make their request and, therefore, the maximum customer wait time is equal to the length of the time slot. Anything more than a small wait time means customers do not experience true VOD. Shorter slot times mean customers will not wait as long which implies a level of service closer to true VOD. However, short time slots also mean that fewer customer requests will arrive during the slot, and that fewer customers will be serviced by one set of resources. Longer slot times mean more customers can be serviced with one set of resources, but the wait time will be longer.

3.2 Providing Interactive VCR Functions

In order to maintain scalability, and create the illusion of one-to-one service, a multicast VOD system must make certain tradeoffs. Because there might be more than one customer in a multicast group, modifying the video stream as a result of one customer's VCR actions in not feasible. Interactivity must be provided through another mechanism. We have developed two ways of providing VCR functions: (1) increasing set-top complexity; and (2) changing the semantics of interactivity.

Increasing Set-Top Box Complexity . Interactive VCR functions can be provided in a multicast VOD system by requiring the set-top box to play the correct frames while still continuing to receive frames multicast by the video server. In most cases, the set-top box can use buffering to handle VCR requests. For example, if a customer wishes to pause a movie, the set-top box can stop playout but continue to receive frames. In the extreme case when the set-top buffer fills up, the customer can be removed from the multicast group and either serviced individually, or added to another existing multicast group. The detailed operation of a system which provides pause-only using set-top box buffering is described in [1].

Changing the Semantics of Customer Interactivity . Interactive VCR functions can also be achieved by modifying the semantics of VCR functions. For example, instead of providing the ability to pause for any duration, a different level of service could offer functions for predetermined time increments. For example, if the base time increment is 2 minutes, a discontinuous pause function would allow a customer to pause for 2,4,6,... minutes. For rewind and fast forward, the customer could move the current playout point by periods of 2,4,6,... minutes in the reverse and forward direction respectively. By making the base time increment equal to the slot time, no additional buffering is needed in the set-top box. A customer initiating a VCR action is removed from the current multicast group, and added to a newly created group, or a previously existing group. A detailed description of a system which provides discontinuous pause, rewind, and fast forward without using set-top box buffering can be found in [2].

4 Simulation and Analysis of a Multicast System

The performance of a VOD system using multicast communication was analyzed by comparing a unicast system; a multicast system with continuous pause; and a multicast system with discontinuous pause, rewind, and fast forward. The performance analysis was based on three measures: (1) *Initial Request Blocking*: the percentage of initial requests that were blocked; (2) *VCR Action Blocking*: the percentage of VCR action attempts that were blocked; and (3) *Undesired VCR Action Length*: the amount of time that a discontinuous VCR action functions beyond what is desired by the customer.

Simulations for all three systems were conducted[1, 2] using a six hour prime-time period. A number of requests were uniformly distributed over the simulation period and for each request a movie was chosen using Zipf's Distribution. Each customer initiated VCR actions according to an exponential distribution, and the duration of each action was uniformly distributed. From the simulations we observed several important results. First, as the number of customers making requests was increased, the initial request blocking probability in the unicast system increased more rapidly. Even for large numbers of customers the initial request blocking probability was acceptable in both multicast systems. Second, as the number of channels was increased, initial request blocking probabilities in the multicast system decreased more rapidly, and approach an acceptable level more quickly. This was true regardless of the level of interactivity provided. Small slot lengths (around 5 minutes) gave good performance and do not significantly affect the on-demand characteristics of the system. As the slot length was increased, the ability of the multicast systems to group and satisfy more requests improved dramatically. The table below summarizes our observations about the three VOD systems. VCR action blocking, which occurs when no resources are available to handle group changes, can be made acceptable in all cases by allocating a sufficient number of emergency channels. Table 1 summarize the numerical results reported in [1, 2].

System	Buffering in the Set-Top Box	On-Demand Nature Of Movies	VCR Actions	Initial Request Blocking	VCR Action Blocking	System Complexity
Unicast w/ Cont VCR	Not Required[1]	True	Continuous	High	Lowest (0%)	Low
Multicast w/ Cont Pause	Required	Near	Continuous Pause Only	Low	Lower	High
Multicast w/ Disc VCR	Not Required[1]	Near	Time Increments	Low	Low	Moderate

Table 1. Results summary of the performance analysis of the three multicast systems.

References

1. K. Almeroth and M. Ammar. Providing a scalable, interactive video-on-demand service using multicast communication. In *ICCCN '94*, San Francisco, CA, Sep 1994. file://ftp.cc.gatech.edu/pub/coc/tech_reports/1994/GIT-CC-94-36.ps.
2. K. Almeroth and M. Ammar. On the performance of a multicast delivery video-on-demand service with discontinuous VCR actions. In *ICC '95*, Seattle, WA, Jun 1995. file://ftp.cc.gatech.edu/pub/coc/tech_reports/1994/GIT-CC-94-49.ps.
3. A. Dan, D. Sitaram, and P. Shahabuddin. Scheduling policies for an on-demand video server with batching. In *ACM Multimedia '94*, San Francisco, CA, Oct 1994.
4. J. Dey, J. Salehi, J. Kurose, and D. Towsley. Providing VCR capabilities in large-scale video servers. In *ACM Multimedia '94*, San Francisco, CA, Oct 1994.
5. T. Little and D. Venkatesh. Prospects for interactive video-on-demand. *IEEE Multimedia*, pages 14–23, Fall 1994.
6. J. Wong and M. Ammar. Analysis of broadcast delivery in a videotex system. *IEEE Transactions on Computers*, pages 863–866, Sep 1985.
7. J. Wong and M. Ammar. Response time performance of videotex systems. *IEEE Journal on Selected Areas in Communications*, pages 1174–1180, Oct 1986.

[1] Minimal buffering is required to handle jitter and provide continuous playout.