Prospects for Interactive Video-on-Demand[1]

T.D.C. Little and D. Venkatesh

Multimedia Communications Laboratory
Department of Electrical and Computer Engineering
Boston University, Boston, Massachusetts 02215, USA
(617) 353-9877, (617) 353-6440 fax
{*tdcl,dinesha*}*@bu.edu*

Abstract–Interactive multimedia systems represent a key technology rapidly evolving from marketing hype and research prototypes to commercial deployments. Supporting truly interactive multimedia services requires providing personalized, dedicated channels or *sessions* for each user. For these services, the user has complete control over the session and has the freedom to explore the depth of both live and stored programming as maintained by future information providers. Development of these interactive multimedia services requires solving a diverse set of technical problems, many of which are tractable in isolation; however, the integration of these many technologies to achieve an end-to-end solution has proven to be a significant obstacle to ubiquitous service.

This paper presents a survey of the technological considerations for the design of a large scale distributed interactive multimedia system. The paper identifies the core problems and examines approaches proposed and implemented for their solution.

Keywords: Video-on-demand, interactive multimedia services, end-to-end system design.

---

1 Introduction

"If you build it, they will interact."[2]

Interactive multimedia technology has seen a great deal of recent attention. If one believes the claims, we will soon be able to view movies, play video games, browse libraries, order pizza, and participate in office meetings (in our bare feet!) from within the confines of our homes. There is also a belief that it represents a technology looking for an application. Appropriately, it has been called an industry that will reach $0 billion per year. This highlights the realization that there is a distinction between the technology (i.e., computers and communications) and the applications that use the technology that will become ubiquitous.

There is good reason for the excitement. Interactive systems will allow their users the flexibility of selecting and receiving specific information using a hybrid paradigm of the TV remote control and information retrieval techniques. This is unlike traditional television delivery mechanisms in which the user is a passive participant receiving what the service provider offers. The availability of enormous communication bandwidth and computing power has made it possible to process and deliver information on a per-user, per *session* basis in real-time. The ability to process information at the source enables the information provider to extract relevant data and modify their characteristics to specific user tastes. Users of such services will have flexibility in choosing the kinds of information they receive. They can also control, based on individual preferences, the manner and time of receipt of this information.

The kind of interaction[3] supported depends to a large extent on the kind of information retrieved and on the particular application under consideration. A home user interacting with movies requires the system to support virtual VCR capabilities (e.g., the ability to play, forward, reverse, and to pause). Most systems use a remote control device (RCD) to browse a menu and select movies on a TV screen. Other interactive features include the ability to avoid or select advertisements, to investigate additional details about news events (e.g., via hypermedia), to save or annotate programming for later reference, and to browse, select, and purchase goods. Clearly, the distinction between the computer and the television is blurring [10].

---

[2]J. Tierney, *New York Times*, June 20, 1993.

[3]Often, service providers distinguish interactive services as *consumption* services in which information delivery is primarily one way (e.g., movies) and *communication* services in which a bidirectional channel is necessary (e.g., teleconferencing).

Several applications have been developed to take advantage of interactive information retrieval and dissemination (e.g., Internet tools like Mosaic and Gopher). Current Internet growth indicates the demand for these services.[4] However, the transmission characteristics of the Internet are not appropriate for high-bandwidth, low latency audio and video-based applications. Telephone and cable TV (CATV) networks have such capabilities and they are establishing their own multiservice networks.

The development of this new information delivery infrastructure will require considerable planning and effort. While most of the media attention has focused on *what* these information systems will deliver, there is little said about *how* these systems will be assembled or what they will cost. Much of the technology required to support these systems exists; however, it is disjoint and significant efforts in integration will be necessary to yield a working end-to-end system. The development of accepted international standards is a first step in the right direction to make this vision a reality (e.g., ATM switching in communication and MPEG standards for video compression). The establishment of standards for set-top devices, video-file servers and metadata databases that enforce interoperability and allow competition among suppliers is important for the growth of such systems. Other issues that require addressing include billing for services, information security, copyright issues, user interface design, and indexing schemes for efficient retrieval.

In the remainder of this paper, we examine the various issues necessary to assemble an interactive information delivery system. Our survey covers the construction of a VOD system from an end-to-end perspective. The paper is organized as follows. In Section 2 we describe interactive services and their implementation. In Section 3 we describe a system architecture for a Video-on-Demand (VOD) system and the technologies required/available for implementation. In Section 4, we perform a study of the problems that must be overcome to make VOD a reality. In Section 5 we survey existing trials underway. Finally, we summarize our views and examine future directions for the technology in Section 6.

2   Interactive Services

Interactive video delivery services represent a fundamental change in the TV interface paradigm. They shift the delivery paradigm from supporting many simultaneous parallel streams (channels) to one that supports concurrent accesses via separate channels into a database. In a traditional broadcast TV system, many stations broadcast their programs

---

[4]There are currently close to 300,000 Mosaic/WWW sites.

simultaneously and the user selects a specific channel to view. As a result, a user is constrained to a chronology of parallel and competing programming. In contrast, an interactive system makes all programming available to its users without this restriction. There is no temporal constraint. All programming becomes available any time to the user.

2.1   Types of Interactive Services

Interactive services can be classified into several categories, based on the amount of interactivity allowed (adapted from [4]).

- **Broadcast (No-VOD)** services that are similar to broadcast TV in which the user is a passive participant and has no control over the session.

- **Pay-Per-View (PPV)** services in which the user signs up and pays for specific programming, similar to existing CATV PPV services.

- **Quasi Video-On-Demand (Q-VOD)** services in which users are grouped based on a threshold of interest. Users can perform rudimentary temporal control activities by switching to a different group.

- **Near Video-On-Demand (N-VOD)** services in which functions like forward and reverse are simulated by transitions in discrete time intervals (on the order of 5 minutes). This capability can be provided by multiple channels with the same programming skewed in time [5, 15].

- **True Video-On-Demand (T-VOD)** services in which the user has complete control over the session presentation. The user has full-function VCR (virtual VCR) capabilities including forward and reverse play, freeze, and random positioning. For T-VOD, only a single channel is necessary; multiple channels become redundant.

PPV and Q-VOD are suitable for services where interaction is limited (e.g., watching movies). In these cases, a local controller ("set-top box") can filter multiple channels to achieve the service. T-VOD requires the use of a bidirectional signal from the user to a centralized controller. T-VOD services are essential for applications supporting multiparty interactive video games, in which instant responses to other players are necessary. Not surprisingly, PPV services are the easiest to implement and T-VOD systems are the most difficult.

Interactive *services* span the range from movies on demand to distance learning (Fig. 1). Table 1 lists some of the basic interactive application areas.

Table 1: Basic Interactive Multimedia Services

Application	Description
Movies-on-Demand	Customers can select and play movies with full VCR capabilities.
Interactive Video Games	Customers can play downloadable computer games without having to buy a physical copy of the game.
Interactive News Television	Newscasts tailored to customer tastes with the ability to see more detail on selected stories. Interactive selection and retrieval.
Catalog Browsing	Customers examine and purchase commercial products.
Distance Learning	Customers subscribe to courses being taught at remote sites. Students tailor courses to individual preferences and time constraints.
Interactive Advertising	Customers respond to advertiser surveys and are rewarded with free services and samples.

From the aforementioned discussion, it is clear that computers, communications, and television are converging. In this context, programming and information can be used synonymously. Through this alliance, some new and interesting features can be added to both environments. In the realm of existing TV use, future functionality will include features such as:

- The ability to choose a source of programming instead of having it forced on the user. (Users are willing to pay extra for this.)

- The ability to skip or select "commercials."

- Virtual VCR capabilities.

- Accessibility to extra information details via hypermedia links.

- A private personal profile used as a filter to screen-out unwanted programming and actively seek out desirable items via agents (knobots).

- Interfaces to view parallel media streams (e.g., video-in-window, subtitles, other events, and multiple views).

- Secure data transfer mechanisms to facilitate electronic catalog shopping and private communications.
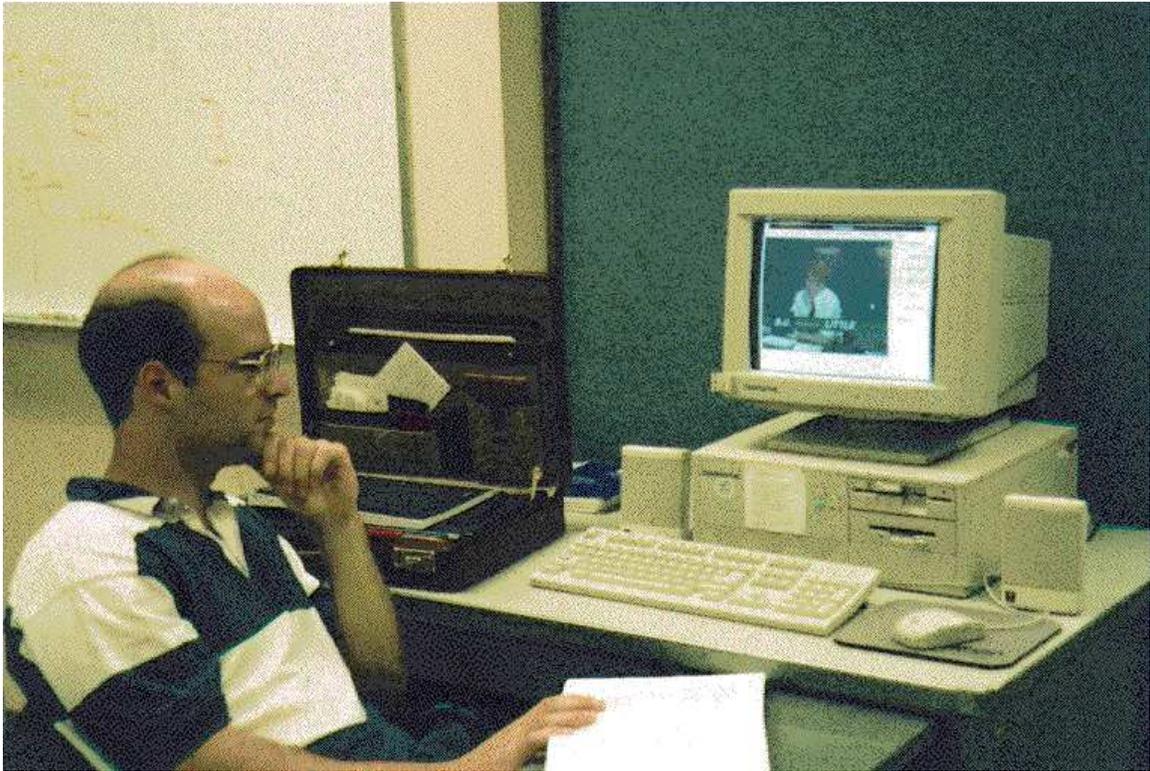
Figure 1: Interactive Distance Learning as a VOD Application

2.2   Technological Inhibitors

Beyond the technology issues involved in the deployment of interactive multimedia applications, there are other inhibitors to ubiquitous deployment. In the all-digital environment characterizing VOD services, information can be readily copied, reproduced, and altered, jeopardizing the established markets of information providers. Before information providers are willing to accept an all-digital system, they need incentives - like mechanisms to protect intellectual property rights - that will maintain their data and thus help them stay in businesses. (*The Internet does not copy data, people copy data.*) Encryption techniques are commonly used in software licensing for operation on limited sets of machines, but these techniques are limited to small distributions. Moreover, software vendors rely on the honesty of their customer and the perishability of their releases. This can also be true of many kinds of video programming (e.g., newscasts).

Digital movie duplication can also be limited through the design of set-top boxes that have limited local storage but allow unlimited re-viewing of programs through storage at the information provider. The ideal mechanism for this purpose would allow flexibility of information delivery and use but would also protect the interests of the information providers.[5]

Another unknown in future VOD systems is the role of advertising. Currently, advertising pays for a significant amount of TV programming. If the aforementioned features are available to the end-user, i.e., the ability to skip the commercials, advertisers may not be so keen on sponsoring the technology. On the other hand, existing trials indicate that users develop new viewing (interaction) habits. This suggests that advertising will have a new place in the realm of interactive multimedia delivery. Remaining is the role of the system designers to develop new techniques to allow users to access "commercials" on demand (COD) so that we do not have to see them repeatedly.

3   System Components for Video-on-Demand

While the individual technologies required to support interactive multimedia services exist and have proven to be viable, there are many architectural design and integration issues that need to be resolved before practical systems become available. A VOD system will

---

[5]A detailed analysis of these issues is beyond the scope of this paper. An interesting survey of the intellectual property rights problem has been provided by Samuelson [14].

be accessed simultaneously by hundreds (if not thousands) of users with different viewing preferences. The system must ensure that the quality of each session remains within specified bounds to achieve customer satisfaction. We survey the individual technologies in the context of an end-to-end architecture for a VOD system.

A typical VOD scenario consists of a local database and server connected to user homes via a communications network. The user home consists of a network interface coupled to a display [4]. The user interacts with the system via a hand-held RCD or a computer keyboard. Fig. 2 illustrates this architecture.
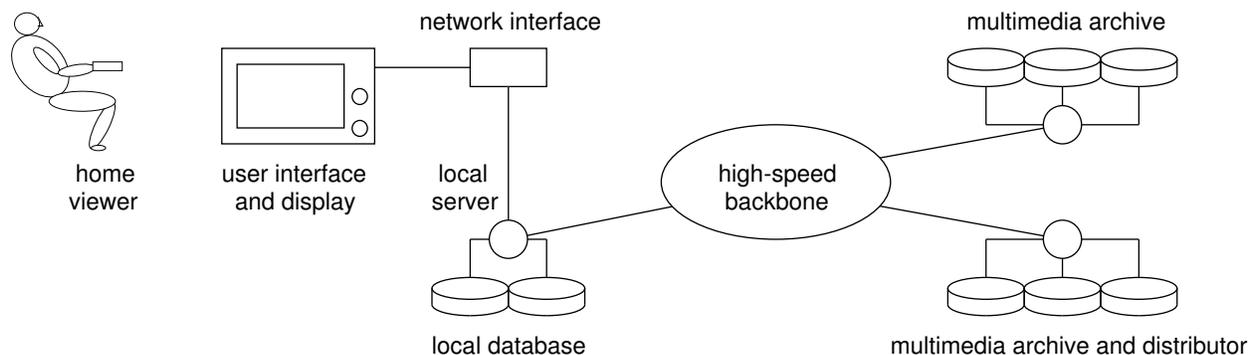

Figure 2: A Simple VOD Architecture

This system consists of an information/programming archive connected via a high speed network to many regional (local) distribution sites. Information is *cached* locally and subsequently delivered to user homes. Such a data distribution scheme serves many purposes. First, it is possible to implement it in a distributed fashion, increasing availability and reliability [6]. Second, one can tailor the information delivery to the specific tastes of a user community in a particular geographic area, reducing costs. Third, it is easier to manage as each local system is responsible for its own billing and accounting. Fourth, the system can be constructed in a regional, piecewise fashion. We will examine the different components that constitute this VOD architecture in greater detail in the following subsections.

3.1 Customer Premises Equipment

To support interactive services, considerable functionality has to be built into the customer premises equipment (CPE). The network interface is responsible for decoding incoming signals and delivering them to the appropriate output device (display and speaker). It also translates user interaction from the remote control to the appropriate signals for network

8

transfer.

The user interacts with the system via a keyboard, RCD, wireless mouse, or voice commands, and a translator (usually an infrared decoder) interprets these commands and forwards them to the network interface for appropriate action. The choice of a hand-held remote system versus a keyboard depends upon the application. For viewing movies, the former is desirable, while the latter are more useful for browsing libraries and distance learning. It is important for the user interface to be simple and easy to use. A schematic of the CPE for home entertainment is shown in Fig. 3 [4].
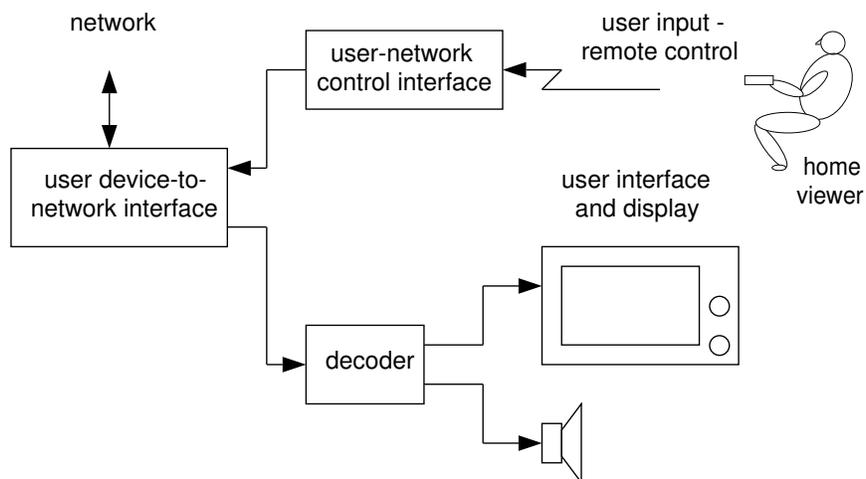

Figure 3: Customer Premises Equipment

The cost of the CPE must be within reasonable limits (under a few hundred dollars) for the technology to be successful. Open and interoperable systems that users can use to subscribe to several different services are preferred [2].

3.2   Access Technologies

Unlike traditional computer systems characterized by short-lived connections that are bursty in nature, VOD sessions are typically long lived (the length of a feature presentation) and require continuous transfer of data. VOD services will require, by today's standards, the delivery of enormous volumes of data to customer homes. For example, entertainment quality NTSC video compressed using the MPEG standards require bandwidths between 1.5 and 6 Mb/s. Many signalling schemes have been developed that can deliver data at this rate to user homes over existing communication links. Some signaling schemes suitable for high-speed video delivery are:

- **ADSL:** The Asymmetrical Digital Subscriber Loop (ADSL) [17] takes advantage of the advances in coding to provide a customer with a downstream 1.536 Mb/s[6] wideband signal, an upstream 16 kb/s control channel and a basic-rate ISDN channel, on existing twisted copper-pair. No additional equipment is necessary if the customer is within 5.5 km of the switching office. The cost to the end user is quite low in this scheme as it requires little change to existing equipment.

- **HDSL:** The High-speed Digital Subscriber Line (HDSL) allows transmission of up to 800 kb/s up to 5.5 km on existing copper lines. With two such circuits in parallel, the technology can support a 1.544 Mb/s full duplex communication.

- **CATV:** This technology uses a broadband coaxial cable system and can support multiple MPEG compressed video streams. CATV has enormous bandwidth capability and can support hundreds of simultaneous connections. Furthermore, as cable is quite widely deployed,[7] the cost of supporting VOD will be significantly lower. However, it requires adaptation to allow bidirectional signalling in the support of interactive services (e.g., T-VOD).

- **SONET:** The SONET (Synchronous Optical Network) is an optical transmission interface standardized by ANSI (American National Standards Institute). The SONET standard specifies the multiplexing format using one or more 51.84 Mb/s channels and defines the optical signal standard for connecting equipment. It also establishes the Operations and Maintenance (OAM) standard and provides a flexible architecture that can provide future Broadband ISDN (B-ISDN) services. The basic signaling rate of 51.84 Mb/s can handle HDTV-quality digital video traffic.

ADSL and CATV are currently the prime candidates for supporting interactive multimedia at the user homes due to their wide deployment. ISDN is a candidate; however, at the basic rate (128 kb/s) it has insufficient bandwidth to support entertainment-quality video delivery.

3.3   Communication Services for VOD

A VOD system requires a high degree of interconnectivity between users and information servers. Moreover, the nature of video traffic requires the transfer of large volumes of data

---

[6]Recently, ADSL-II, which can deliver data at the rate of 6 Mb/s into user homes has been proposed.

[7]Cable has been laid past approximately 65% of all homes with a penetration of about 90% in the United States [10].

at very high speeds. Many communication protocols and network architectures have been proposed to connect the various components including (i) the Asynchronous Transfer Mode (ATM), (ii) Fiber Distributed Data Interface (FDDI), (iii) DQDB (Distributed Queue Dual Bus), and (iv) the 100 Mb/s Ethernet (IEEE 802.12). Of these proposals, ATM is emerging as the frontrunner for implementing VOD systems. ATM combines the advantage of packet and circuit switching schemes. It supports resource reservation and virtual circuit connections over packet switched networks, allowing guaranteed service and interoperability. In addition, many networking vendors have expressed support for this technology and ATM products are already available. FDDI, DQDB, and the 100 Mb/s Ethernet are suitable for communication within a small office or business environment where resources need to be shared and a very high bandwidth is not necessary. For the academic Internet, the IETF task force is working on developing a TCP/IP interface for ATM.

Currently, ATM switching is being installed in the backbone networks in the United States. Several long distance carriers and RBOCs have begun to offer ATM services to their customers. As demand grows, more hubs will be installed closer to user homes.

3.4   Storage Servers

To support a large user population and a wide variety of information/programming, a VOD system must store and manage large volumes of data. A 90 minute feature presentation, using MPEG-I compression, requires about 1 Gbyte of storage. Table 2 illustrates some basic storage economics and provides a rough estimate of the cost for different storage device types.

It is apparent that a larger number of storage devices yield many independent sessions as measured in available bandwidth. A larger number of T-VOD sessions and improved user interaction capabilities can be provided by achieving aggregate storage capacity with many devices. However, this basic analysis assumes a perfect mapping of user movie choices to available storage, i.e., a selected movie can be assigned to the free I/O bandwidth of any system. In reality, user movie choices must be mapped to sessions from devices that store the selected movies. If an imperfect mapping is used, the limitations in I/O bandwidth prevent this perfect utilization. These topics are explored further in Section 4.

Solid-state storage technologies such as Flash memory and RAM are currently too expensive to store the terabytes of data required for VOD services. However, they can support more concurrent interactive sessions compared to disks. While tape memory is cheap, the extra swap and seek times they introduce do not make them attractive from the perspective of

11

Table 2: Basic Storage Economics

Storage Type	Cost/MB	Sessions/device	Cost/Movie Stored
RAM	$ 50	200	$ 50,000
Hard Disk	$ 0.5	5	$ 500
R/W Optical	$ 0.2	2	$ 200
Magnetic Tape	$ 0.01	1	$ 10

providing interactive services. A hierarchical storage scheme offers an attractive means of satisfying the memory bandwidth needs of VOD systems [3]. In the hierarchical scheme, the more popular movies are stored on RAM, the less popular ones on hard disk, and the least popular on tertiary storage. Optical (CD-ROM, WORM, etc.) and magnetic (tape drives, disks, RAID, etc.) systems will be sources of inexpensive tertiary storage for archival purposes. The hierarchical approach also yields reduced operating costs and can offer a wide selection to its users.

There are many reasons why the downloading and caching of entire movies at user homes is not being considered by most current VOD system designs:

- Often, communication bandwidth is limited at user homes and can support a rate just enough for a single media stream. To cache information, higher data rates are necessary. Otherwise the user must wait for most of the information to arrive. This delay is unacceptable for most scenarios.

- Caching an entire movie requires a large amount of memory (> 1 Gbyte) at the user home and will increase the cost of the end-system [9].

- Information providers are unlikely to be willing to allow an entire program to be available at once as this provides some protection from data duplication and piracy.

However, if the cost vs. performance factor increases in favor of solid-state memory systems in the future, we foresee their increased usage in the set-top box for caching information.

3.5   Schemes for Query Support and Information Retrieval

While it is interesting to describe the realm of information that interactive technology will bring to its users, the issue of management and presentation of data remains. It is

impractical to expect a user to browse 500 channels to select a desired program. This process is currently simplified by an indexing mechanism in the form of a printed listing of the day's programming represented as a chronology. Because future systems will be moving to a combination of many channels and stored information sources, new means for locating, browsing, and retrieving programming are necessary [13, 7]. These mechanisms will index and cross reference information and employ hypermedia techniques to simplify the search process. Application-specific models are necessary to support application functionality (e.g., searches on movie scene content). The task of creating indices for video is difficult due to the unstructured nature of information [16, 1]. This leads us to the realization that there exists a need to extend image content queries to the video itself. However, a detailed discussion of this topic is outside the scope of this paper.

For faster access times, a database system can use a "metadata" approach for information management. Metadata are "data about data." A metadata server contains concise information about the location and characteristics of the data to be retrieved (e.g., movie titles, where they are stored). By using such a scheme, the data delivery process (raw video communications) can be decoupled from the database management functions.

The advantage of the metadata approach is that it allows the user to examine the contents of a database without having to retrieve the large data objects. The user can examine the information summary and select the programming to be delivered. This concept is particularly appealing with respect to stored audio and video that can be costly to deliver and can consume significant portions of a data server's I/O bandwidth. The requirements of metadata management to support interactive multimedia applications include basic services to support object identification, location, naming, and distribution.

The metadata concept is ideal for interactive multimedia applications in which many objects can be involved in a single presentation. For example, a movie might involve the coordination of audio and video streams. A classroom video might include the simultaneous presentation of audio, video, graphics and text. The metadata server maintains the relationships between these diverse objects and ensures their proper delivery to the user home.

Another attraction of using metadata servers is the flexibility they offer. The metadata server can be coupled to the resource management mechanism (resource server) to better manage system resources. When resource updates occur (resource removal, relocation, and new resource generation), only the entries in the database need updating for the change to be visible to the entire user community. Finally, the metadata server itself can be replicated, providing for an additional level of performance.

For a VOD system, the process by which sessions are established can be viewed as a combination of two processes. (i) a *Location Phase* when the location of the data to be retrieved in the network is determined (via the metadata server), and (ii) a *Setup Phase* when the communication parameters are negotiated and data delivery begins. In most existing computer systems, the location and setup phases occur independently (e.g., Internet Archie, Gopher, WAIS, etc.). The user determines the location in the network of the node containing a desired data using one service, and sets up the connection using another. No information is available to the user regarding the availability/nonavailability of the resource. The user can get tied down searching for specific information. Such delays can be especially critical in VOD systems, where an immediate response is preferred. The above limitation is not true for telephone systems where the routing mechanism is transparent to the user.

4   Management of System Resources in VOD

In the previous sections, we identified some of the technical problems in designing a VOD system. A VOD system must support a large customer population and many movie titles. Most existing prototypes are confined to laboratory or office environments and support at most a few hundred users and up to a hundred movies. Large scale commercial systems must need to more closely match the per-user resource requirements and usage patterns to achieve economic feasibility. In this section, we address some of these problems and discuss existing research in this area.

4.1   Resource Reservation

A fundamental problem in developing a VOD system is one of storage and network I/O bandwidth management. The VOD system has a finite amount of resources measured in terms of storage I/O and communication bandwidths. As customers compete for the same system resources, efficient schemes that ensure fairness of allocation have to be designed. The service provider desires to generate the maximum revenue from the offered services. A balance between these two often conflicting requirements is necessary to tap the potential benefits of the system. The first step towards solving this problem is the development of an accurate system model. We use the model proposed in Fig 2 as the basis for the remainder of this discussion.

The end-to-end VOD system consists of three basic components; the *storage server*,

the *network*, and the *user interface.* The *metadata server* adds an additional level of complexity to the system model. The time dependency of continuous media requires the VOD system to ensure that the data transmission mechanism can support strict deadlines. If these deadlines are missed, the quality of the session can degrade. To ensure customer satisfaction, resources must be reserved along the entire data path of a connection on a per-session basis. The complexity of the resource reservation mechanism depends upon the application under consideration. For example, in broadcast services only network resources have to be reserved as all destinations receive the same data. In contrast, interactive services require that resource reservation be made per-session along the entire data path, including at the source.

A crucial factor affecting resource reservation is Quality-of-Service (QOS). Unfortunately, the common interpretation of QOS is from a network perspective rather than a user or customer perspective. A more suitable view decouples the two perspectives and yields two QOS characterizations (we can call them delivery quality and system QOS). A current challenge is to identify the mapping from delivery quality to system QOS for a range of system design parameters (e.g., data compression and network switching modes).

4.2   User Traffic Characterization

Even though customers access the VOD system randomly, having *a priori* knowledge about user access patterns can lead to a more efficient design. The system can use this information to manage network and storage bandwidths. For example, if the traffic characteristics indicate that a movie is popular at a particular site, the system can replicate the movie locally to increase availability.

Typically, the behavior of a single user accessing the movie database can be described as follows: When a user accesses a movie from the database, he or she might decide that the movie is not interesting after a few minutes of viewing and decide to browse for a different selection (this is similar to *channel surfing*). On the other hand, if the user has already watched a considerable portion of the movie, the likelihood of termination before movie completion diminishes with time. This information can be used to commit resources to the user only after the movie has been viewed for a threshold time. This approach can be advantageous in schemes where the data is prefetched and cached at the user home.

Similarly, the access pattern of users to the system will not be uniform over a given 24 hour period. Typically, one would expect the load to be low to moderate during the

daytime and to increase gradually through the evening and fall off again during the night. A hypothetical schematic characterizing the access to a VOD database for a 24 hour period is shown in Fig. 4. The access to the database is high during the evening hours, peaks at around 9:00 PM, and is low-to-moderate during the day. This access pattern can be used for designing schemes for resource management; to update popularity tables, redistribute data, and reconfigure the system during off-peak hours.
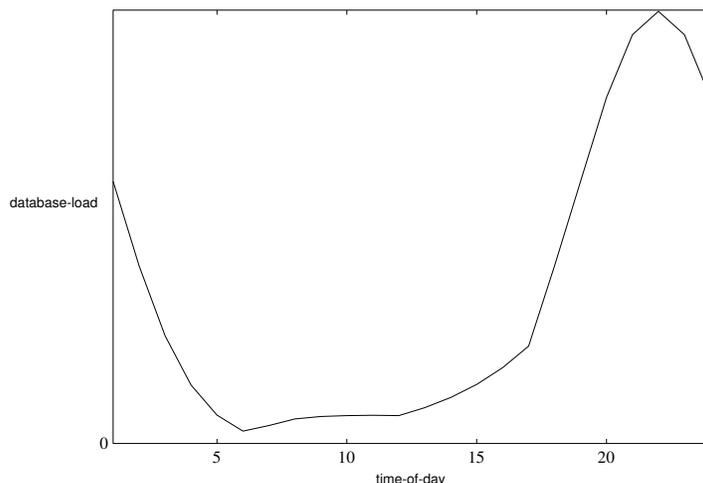

Figure 4: A Schematic Daily-Access Model for a VOD System

Similar models can be developed and maintained for different geographical regions, movie categories, and individual titles. Such models can accommodate the differences in programming choices (e.g., children's movies are more popular during the early evening hours) of different user groups. However, the complexity of these models, and their tractability is yet to be established.

4.3   Load Balancing

An issue directly related to resource reservation is load balancing. In the context of VOD, load balancing can be viewed as a combination of two sub-problems (i) The movie-storage allocation problem and (ii) the resource location and connection establishment mechanism. Even though these problems are more easily solved individually, they are not independent with respect to performance. Solving these issues from the perspective of a generic interactive system is an open problem; however, simplifications can yield tractable solutions.

For example, if one assumes that a VOD system supports only stored data; i.e., movies have to be digitized and stored before they can become available online, then the data

characteristics of a movie are well known in advance (e.g., the system has *a priori* knowledge about the average bandwidth, burst rates, burst durations, etc.). This knowledge can be used to simplify the design process. Furthermore, the nature of VOD leads us to believe that the network protocols required to support sessions are connection-oriented rather than connectionless. Telephone systems support similar connection-oriented sessions and we believe that much of the existing work in telephony can be used for developing resource allocation and location mechanisms. Using the metadata mechanism as described in Section 3 simplifies the task of management by decoupling the storage problem from the location problem.

4.3.1 Movie-Storage Data Distribution

Using a VOD architectural model as illustrated in Fig. 2, a reasonable movie-disk distribution scheme can be performed as follows. On a daily basis a centralized movie distributor calculates the current movie popularity. From these popularities movies are replicated as necessary and distributed during off-peak hours to localized storage systems. The movies are then available for viewing during peak hours based on anticipated demand [6, 9].

Although centralized control is assumed for movie redistribution, the archival of movie "originals" can be done in either a distributed or centralized fashion. One can view the archive as a very large database containing all available movies and stored on media with varying access latencies and I/O bandwidths. This model is not unlike the multi-tiered memory architecture of computer systems. Rare requests for information not available locally will be satisfied by transferring information at high speeds from the archive and caching them at the local server for playout.

Other interesting problems related to resource reservation and management include:

- **QOS Negotiation:** The connection establishment procedure between the user's machine and the video server requires the negotiation of QOS parameters for the session. Schemes that achieve QOS negotiation and session establishment transparently are necessary.

- **Scalable Services:** Scalability is a broad term used to encompass many service types. In context of VOD, it is applicable to the server's ability to support the data requirements of multiple terminal types (e.g., color vs. monochrome, HDTV vs. NTSC). Scalability is also relevant in the context of service quality. Service providers

can reduce interactivity during high load periods to accommodate more sessions, or charge a higher fee for the same service.[8] The effect of these schemes on user behavior is not immediately apparent.

- **Advance Reservation:** Users can request and reserve resources in advance. This approach makes the job of resource allocation easier as the service provider has some *a priori* knowledge about the projected loads. Service providers can make advance reservation attractive to customers by providing discount rates for early reservations as is common for airlines and conferences.

5 Existing Trials

Many trials are currently underway, both in the commercial and academic worlds to test the feasibility of providing interactive services. Table 3 lists some commercial trials currently underway.

Table 3: Some Commercial Trials of Interactive Services

Location	Company	Description
Orlando, FL	Time Warner	Limited VOD to homes via cable
Cerritos, CA	GTE	Primitive VOD Trial using VCR's
Montreal, Canada	Videoway	Quasi-VOD via cable
Northern Virginia	Bell Atlantic	VOD using ADSL
New York	NYNEX	VOD Trials in Manhattan
New Hampshire/Cambridge	Continental Cable	Connection to Compuserv via Cable
California	Sega	Interactive Video Games
Boston, MA	GTE	"Main Street" with 50 services

Notable research prototypes include those of Miller et al. [8] who describe a prototype system developed at Dow-Jones in collaboration with NYNEX for the delivery of business news. Other systems in this area are the Berkeley CM Player[12], and the Virtual Video Browser (Fig. 5)[7]. However, all these prototypes are small scale designs and support at most a few tens of users. Relevant information access systems that provide interfaces to the Internet include Mosaic, Archie, and Gopher. These tools make information access easy on a large scale, but are primitive in their support of continuous media (audio and video). European trials include the PAEN (Pan European ATM Network), GEN (Global European Network), and RACE in Europe.

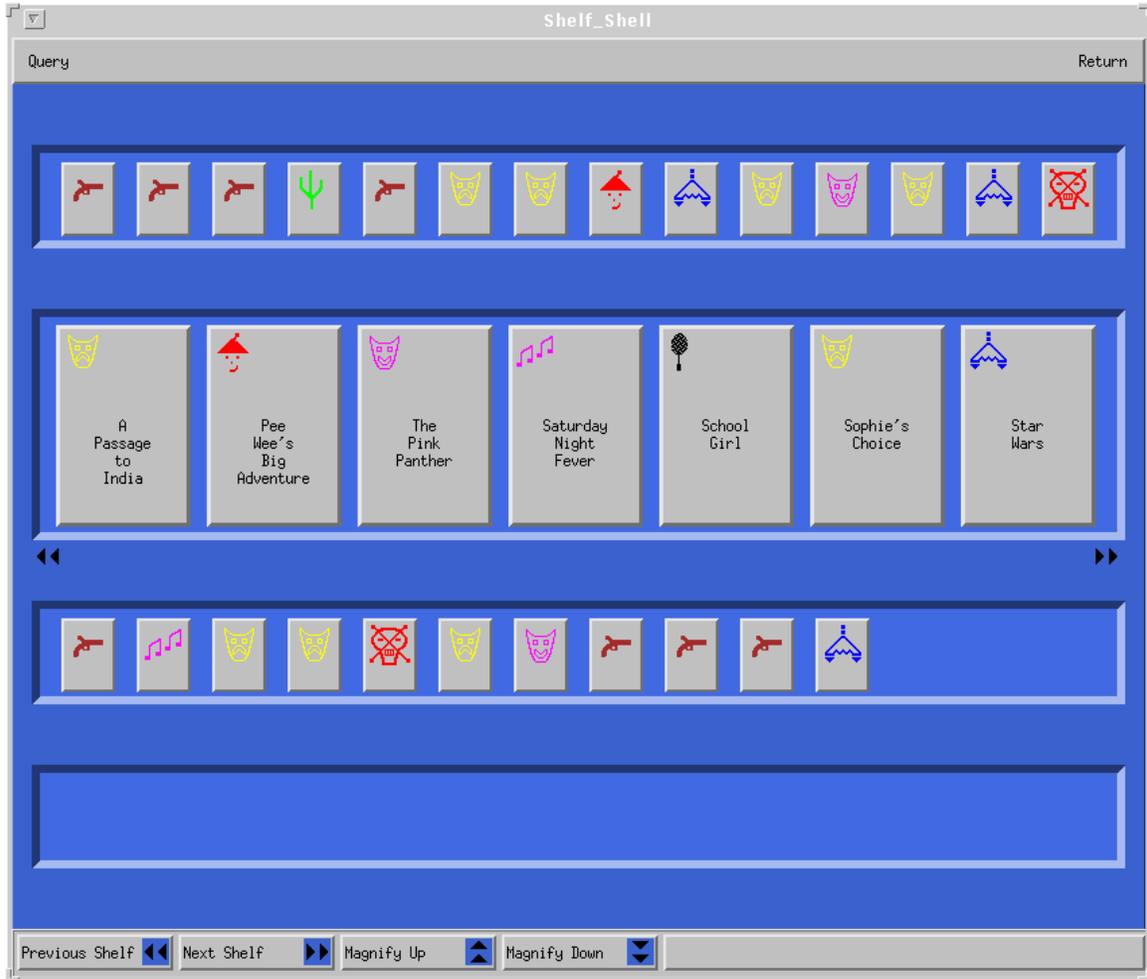[8]This is analogous to the telephone system which has reduced rates during the night.

Figure 5: A User Interface for Interactive Movie Delivery

Oracle Corporation is developing a massively parallel server based around a nCube processor to support multiple concurrent sessions. IBM has also developed a massively parallel server called the SP2. Other Corporations (SGI, HP, Sun, DEC, Data General) have developed multiprocessor servers. Other companies (e.g., Apple, Microsoft) are racing to develop the GUI interfaces for these systems. Several other organizations (General Instruments, SGI, Nintendo) are actively designing customer set-top boxes. All this activity leads us to the conclusion that it will not be long before interactive services become as prevalent as phone and television services.

6 Summary

In this paper we have examined a number of issues relating to the deployment of large scale interactive VOD systems. To become a viable alternative to existing information delivery services VOD systems must support many users and movie selections at affordable rates. Most of the existing prototypes in this area are very primitive in their ability to support such large levels of interaction. While much of the technology required to make VOD a reality exists, integrating the different pieces into a working end-to-end system will require significant planning and effort. The VOD system will require efficient schemes to allow indexing and browsing along with the ability to handle the transfer and storage of large volumes of data.

The transition to truly interactive systems will be a gradual one, with existing systems evolving to support different applications and features. The deployment of VOD services to all user homes will be very expensive and initially, services will become available over existing infrastructure. As this paper is written, new features and systems are entering the marketplace, new regulations are being proposed, and industrial alliances are being formed and broken. The future will be an exciting one.

Acknowledgments

References

[1] F. Arman, A. Hsu, and M.-Y. Chiu, "Image Processing on Compressed Data for Large Video Databases," *Proc. ACM Multimedia*, Anaheim, CA, 1994, pp. 267-272.

[2] Y.-H. Chang, D. Coggins, D. Pitt, D. Skellern, M. Thapar, and C. Venkataraman, "An Open-Systems Approach to Video on Demand," *IEEE Communications*, Vol. 32, No. 5, May 1994, pp. 68-80.

[3] C. Federighi and L.A. Rowe, "A Distributed Hierarchical Storage Manager for a Video-on-Demand System," *Storage and Retrieval for Image and Video Databases II, IS&T/SPIE Symposium on Elec. Imaging Sci. & Tech.,* San Jose, CA, February 1994.

[4] A.D. Gelman, H. Kobrinski, L.S. Smoot, S.B. Weinstein, M. Fortier, and D. Lemay, "A Store and Forward Architecture for Video On Demand Service," *Proc. IEEE ICC*, 1991, pp. 27.3.1 - 27.3.5.

[5] W. Hodge, S. Martin, and J.T. Powers, Jr., "Video On Demand: Architectures, Systems and Applications," *SMPTE Journal*, September, 1993, pp. 791-803.

[6] T.D.C. Little and D. Venkatesh, "Probabilistic Assignment of Movies to Storage Devices in a Video-On-Demand System," *To appear in ACM/Springer Multimedia Systems*, 1994.

[7] T.D.C. Little, G. Ahanger, R. Folz, J.F. Gibbon, F. Reeve, D. Schelling, and D. Venkatesh, "A Video-on-Demand System Supporting Content Based Queries," *Proc. ACM Multimedia*, San Diego, 1993, pp. 427-436.

[8] G. Miller, G. Barber, and M. Gilliland, "News On-Demand for Multimedia Networks," *Proc. ACM Multimedia*, San Diego, August 1993, pp. 383-392.

[9] C. Papadimitriou, S. Ramanathan, and P.V. Rangan, "Information Caching for Delivery of Personalized Video Programs on Home Entertainment Channels," *Proceedings of the IEEE International Conference on Multimedia Computing and Systems,* Boston, 1994, pp. 214-223.

[10] L. Press, "The Internet and Interactive Television," *Communications of the ACM*, December 1993.

[11] S. Ramanathan and P.V. Rangan, "Architectures for Personalized Multimedia," *IEEE Multimedia*, Vol. 1, No. 1, Spring 1994, pp. 37-46.

[12] L.A. Rowe and B.C. Smith, "A Continuous Media Player," *Proceedings of the 3rd International Workshop on Network and Operating System Support for Digital Audio and Video,* San Diego, November 1992.

[13] L.A. Rowe, J.S. Boreczky, and C.A. Eads, "Indexes for User Access to Large Video Databases," *Storage and Retrieval for Image and Video Databases II, IS&T/SPIE, Symp. on Elec. Imaging Sci. & Tech.*, San Jose, CA, February 1994.

[14] P. Samuelson, "Copyright's Fair Use Doctrine and Digital Data," *Communications of the ACM*, Vol 37, No. 1, January 1994, pp. 21-27.

[15] W.D. Sincoskie, "System Architecture for a Large Scale Video on Demand Service," *Computer Networks and ISDN Systems*, Vol. 22, 1991, pp. 155-162.

[16] S.W. Smoliar and H.-J. Zhang, "Content Based Video Indexing and Retrieval," *IEEE Multimedia*, Vol. 1, No. 2, Summer 1994, pp. 62-72.

[17] J. Sutherland and L. Literal, "Residential Video Services," *IEEE Communications Magazine,* July 1992, pp. 36-41.