Adaptive Large-scale Multimedia Storage Systems

Leana Golubchik

Abstract

This proposal requests funds for release from teaching duties during the Fall 1998 semester, in order to devote full time to a research project for design and development of adaptive large-scale multimedia storage systems (i.e., systems that support such applications as Video-on-Demand (VOD), digital libraries, etc.). Existing research funding is currently being used for partial summer salary support of the PI; it is also intended for partial support of a graduate research assistant, in the future. The semester for which GRB support is requested, Fall 1998, will be critical for establishing a multimedia systems laboratory as well as for initiating graduate student research projects.

The broad aim of our research is to develop a novel class of multimedia storage technologies that improves access by several orders of magnitude. The key idea that we pursue is to exploit the intrinsic characteristics of multimedia applications in organizing and scheduling storage access. Thus, we investigate media-aware, workload-aware, and application-aware storage technologies that are optimized to support multimedia applications access. Specifically, in this proposal, we will investigate: (a) the use of *selective media retrieval and delivery* schemes, which will allow dynamic adjustment of system behavior to the current availability of resources and workload demands and (b) efficient *resource management* techniques through the use of *predictive methods* in the face of interactive workloads.

Many systems today are moving towards applications whose primary function is to transfer large volumes of data from a storage subsystem through the network (possibly subject to some form of real-time constraints). Thus, we expect the impact of our research to be a new generation of storage system architectures and novel classes of resource management algorithms that will be applicable across many classes of large scale systems.

1 Goals

The broad aim of our research is to develop a novel class of multimedia storage technologies that improves access by several orders of magnitude. The key idea that we pursue is to exploit the intrinsic characteristics of multimedia applications in organizing and scheduling storage access. There undoubtedly exists a multitude of exciting, promising, and much needed multimedia applications. Quite a few of these promising applications are beginning to emerge and impact our lives, however, most exist on a "small scale" at the moment. As we begin to overcome the technological difficulties in image processing, computational power, signal processing etc. required by multimedia applications, they begin to grow to their life size potential, i.e., terabytes of video, audio, textual, and image data. The designs of storage systems that succeed for "small" versions of multimedia applications simply do not scale-up[1]. Therefore, without properly architectured storage systems, we will not be able to build "large scale" versions of these applications and fulfill the promise of multimedia information systems that can significantly alter the way science, business, education, entertainment, and even the mundane aspects of our lives are conducted.

In designing and building large high performance multimedia information systems, one must consider a whole spectrum of applications, from relatively low bandwidth, high throughput, and "just-in-time" delivery of Video-on-Demand (VOD) servers to very high bandwidth, relatively low volume, and "ASAP" delivery of supercomputing/scientific applications. Such systems must be able to accommodate the various storage, performance, and reliability requirements of the different types of media and applications, often coupled with real-time constraints which are due to the isochronous nature of media such as video and audio. Hence, scalability, real-time data delivery, and consideration of the entire pool of system resources and how they interact have been the underlying themes of our work, which thus far has clearly illustrated, at least in the context of VOD systems, that proper resource management can lead to significant performance improvements and significant increases in the size of problems and applications that can be supported by a particular architectural configuration. We continue our work in large multimedia information systems, in the context of a more general class of multimedia applications. More specifically, in this proposal we will concentrate on the following approaches to media-aware, workload-aware, and application-aware storage system design:

- Use of *selective media retrieval and delivery*.
- Efficient *resource management* through the use of *predictive methods*.

We elaborate on these approaches in Section 3.

[1] Commercial tests of applications such as Video-on-Demand systems, conducted in the last few years, indicate that storage server designs are difficult to scale-up when attempting to service thousands of users simultaneously.

2 Work Completed to Date

There has been a lot of work done in the area of multimedia storage servers in the last few years, and part of the effort has been simply to understand what is possible. Thus, one of the goals of our work thus far has been to illustrate what the tradeoffs are and how to go about making proper design choices when building such systems. The issues that must be addressed in the delivery of multimedia data include: (a) data layout, (b) scheduling, (c) access control, (d) fault tolerance, and (e) management of the storage hierarchy, where the resources include: (1) I/O (e.g., disk) bandwidth, (2) storage space, and (3) buffer space. The amount of resources plus the data layout, scheduling, etc. schemes that are used will determine the *quality of service* (QoS) that can be provided by a particular system. For instance, we can view QoS in terms of *latency*, *throughput*, *reliability*, etc. characteristics of a system. Below, we briefly discuss several results of our work to illustrate that through proper resource management we can design low latency, high throughput, and highly reliable systems.

To exhibit reasonable performance, multimedia storage servers require large disk farms, which have a high probability of (a single) disk failure. Due to the real-time constraints, the reliability and availability requirements of multimedia systems are even more stringent than those of traditional information systems. In [2, 1] we introduce the first non-RAID-based fault-tolerant design of a multidisk VOD storage server. This research has: (a) shown that RAID-based solutions (often used for providing fault tolerance in traditional disk-based systems) are inadequate for VOD servers due to poor resource usage and (b) produced data layout and scheduling techniques which resulted in highly reliable but significantly more cost-effective video servers with higher throughputs and lower latencies (than more traditional, previously pursued RAID-based techniques). Our current work on this topic concentrates on reduction of degradation in system performance under failure and rapid, but "non-performance-degrading", recovery from failure to normal operation.

As *storage* costs continue to decrease rapidly, we expect disk bandwidth to become one of the limiting resources in multimedia storage servers, contributing to high latency and low throughput. In [3, 4] we introduce a novel data sharing method, termed "adaptive piggybacking", for reducing the (aggregate) disk bandwidth requirements of a VOD server which accomplishes significant throughput improvements while eliminating the cost of additional latency or additional resources (incurred by previous approaches to data sharing, namely batching and buffering). This work includes development and analysis of various policies for performing "adaptive piggybacking". Our current work [5] on this topic concentrates on: (a) combining the various data sharing techniques in designing more cost-effective media servers (these techniques are not mutually exclusive) as well as on (b) constructing performance models for proper resource allocation and system sizing in media servers exploiting data sharing techniques.

The research proposed here aims to build on this work, for instance, by extending our techniques to a more general class of applications (than VOD servers). The details of the specific tasks that we will pursue to this end in this proposal are given in Section 3.

3 Calendar of Research Activities

This project consists of investigation of the following areas. Firstly, we propose to investigate the use of *selective media retrieval and delivery* techniques, which will allow dynamic adjustment of system behavior to the current availability of resources and workload demands. More specifically, this will include development of methods for providing continuity in delivery of multimedia information in the face of fluctuations in: (a) storage system workload, (b) network congestion, as well as (c) availability of system resources. Secondly, we propose to investigate efficient *resource management* through the use of *predictive methods*. This will include design of *resource sharing* techniques, specifically through the use of *predictions* for near-future data requests in the face of changes in workload due to the interactive behavior of multimedia applications. Finally, the project will include an experimentation phase where we will evaluate the usefulness of above methods.

All research will be conducted at the University of Maryland; we will be working with computing equipment provided by the University of Maryland Computer Science Department. We expect to complete the experimental (laboratory) set-up by approximately half-way through the project and begin experimentation and evaluation at that point.

Specifically, our research activities will focus on the following tasks, each taking about $\frac{1}{3}$ of a semester: (1) [1.5 months] development of selective media retrieval and delivery techniques, (2) [1.5 months] development of efficient resource management techniques through the use of predictive methods, and (3) [1.5 months] experimentation and evaluation of methods developed under the first two tasks. We briefly describe each one of these tasks in more detail below.

3.1 Selective Media Retrieval and Delivery

As a user moves about in a mobile computing environment, or as new users arrive to a more "traditional" multimedia information system, there is a need to dynamically adjust delivery of multimedia information to the changes in the system's workload. Since many requests being serviced by a multimedia storage server exhibit long service times (due to the size of data and nature of the applications) and given QoS (e.g., real-time) constraints, the dynamic adjustments in delivery of data in such systems are significantly more difficult. Our goal is to provide continuity in delivery of multimedia information in the face of fluctuations in: (a) storage system workload, (b) network congestion, as well as (c) availability of system resources. For instance, the changes in *workload* can occur due to changes in the *rate* at which information is being accessed (e.g., as a result of a fast-forward request in a VOD server) or due to sudden changes in *what* information is being accessed (e.g., a user walking through a virtual office building suddenly runs outdoors, thus destroying any form of gradual change in the images that must be retrieved and displayed). On the other hand, the *resource availability* can change, for instance, due to failures (either of storage components or communication links) or due to mobility characteristics of the environment (e.g., a user entering a new cell (or "communication location") may not find the resources to continue receiving data at the same bandwidth as in the old cell). Note that these changes in resource *usage* and *availability* can occur both on the storage server as well as in the communication network, and the storage server must be able to react to the changes occurring in the *communication* network just as well as to those occurring in the storage system.

Thus, we propose to investigate techniques for *selective* multimedia data retrieval and delivery, which will allow dynamic adjustment of system behavior to the current availability of resources and workload demands. In our earlier work, we have investigated "selective" data retrieval and delivery in the face of disk failures [2] and proposed schemes which can significantly increase the number of simultaneously serviced streams in a VOD server. Our current directions [7], which we will pursue in this proposal, include investigation of support of different levels of QoS by exploiting: (a) the multiresolution property of video (and images) and (b) redundant information stored in the system for fault tolerance purposes. The multiresolution property would allow us to store video (and images) in a layered fashion, where (basically) the use of each additional layer will improve the quality of delivered data. The redundant information would allow us to have "choices" at data retrieval time and thus facilitate system load balancing.

The focus of our work will be on scheduling of data retrieval and approaches to resolving disk bandwidth congestion under expected and unexpected changes in resource demand. Specifically, we consider will techniques which dynamically adjust resolution of video streams in progress in order to adjust to fluctuations in workload while satisfying given QoS constraints and utilizing system resources efficiently. In this proposal we will investigate resolution adjustment and load balancing techniques which address: (a) different causes of workload fluctuation, (b) the extent and duration of "overflow" of resource demand beyond the available amount, (c) predictability of future resource demand, and (d) variations in QoS requirements.

3.2 Efficient Resource Management through Predictive Methods

Interactive applications, for instance, such as virtual world environments, where users can control to some extent what is visible, what will be visible next, how soon the changes will occur, and so on, present a whole new world of challenges to multimedia storage servers. Part of the difficulty is in attempting to construct timely and efficient data retrieval and delivery plans in the face of frequent changes in workload due to the interactive nature of these applications. Such scheduling problems can be significantly more difficult than the corresponding data layout and scheduling problems that have thus far been studied in the context of VOD servers, where the interactivity is primarily due to VCR functionality requests (e.g., fast-forward, rewind, etc.).

Our goal in this proposal is to design *resource sharing* techniques, specifically through the use of *predictions* for near-future data requests in the face of changes in workload due to the interactive behavior of applications. In our earlier work [3], we have shown that the use of data sharing techniques in the context of VOD servers can significantly improve system performance. Our current directions [6], which we intend to pursue in this proposal, include investigation of techniques for reducing the overall load on the storage subsystem by partitioning a set of client requests into groups, based on predictions for their future data demands, and servicing each group with a single set of resources. The attempt here is to construct a somewhat more general (or unifying) framework for accomplishing this goal that can accommodate: (a) a large set of techniques for sharing resources among requests in a group and (b) a large set of applications, especially those with interactive environments.

3.3 Experimentation and Evaluation

The last task in this proposal involves experimentation with and evaluation of techniques developed under the two tasks described in Sections 3.1 and 3.2 above. We intend to implement the techniques developed under these tasks and evaluate them, based on criteria such as: (a) QoS provided, where examples of QoS measures include jitter, percent of data loss, etc. and (b) cost/request, e.g., we will consider the maximum throughput that the system can support

and the cost of the system architecture needed to support it — different resource management techniques will result in savings of different resources, and since it is not immediately clear how to compare savings in one resource versus another, one approach is to assess the tradeoff through cost considerations. It is also worth noting that different resource management techniques have their optimum operating points at different architectural configurations; thus it is usually not the case that one resource management scheme is absolutely better than another, but rather that one must understand the system requirements and constraints and then choose a resource management scheme accordingly.

4 Final Form of Project and Significance of Results

In summary, there is a multitude of real-time and interactive applications of multimedia storage servers, including video-on-demand, scientific visualization, simulated world environments, and many others. Technological advances in digital signal processing, data compression techniques, and high speed communication networks have made such systems *feasible*. Through efficient use of resources in properly architectured multimedia storage systems we can make these applications *accessible*.

Our work is focused on design of cost-effective and scalable mechanisms for real-time delivery of multimedia data, in the context of storage subsystems, while cognizant of similar issues in related areas, such as communication networks. Many systems today are moving towards applications whose primary function is to transfer large volumes of data from a storage subsystem through the network (possibly subject to some form of real-time constraints). Thus, we expect the impact of our research to be a new generation of storage system architectures and novel classes of resource management algorithms that will be applicable across many classes of large scale systems.

We expect to submit our results to leading conferences in the field, such as SIGMETRICS, SIGMOD, VLDB, etc. as well as leading journals in the fields, such as IEEE Transactions on Knowledge and Data Engineering, IEEE Transactions on Parallel and Distributed Systems, ACM Multimedia Systems Journal, etc. One of the benefits derived from this project will be an implementation of a multimedia storage system which we intend to use for both research and educational purposes. One advantage of having such a system operational and in-use is the ability to collect workload information, which can prove to be critical in making proper design choices (little information exists on this topic currently). Of course, another motivation for having such a system in place is to use it for testing and improving our ideas.

References

[1] S. Berson, L. Golubchik, and R. R. Muntz. Design of High Performability Low-Cost VOD Servers. *Submitted to IEEE Transactions on Parallel and Distributed Systems.*

[2] S. Berson, L. Golubchik, and R. R. Muntz. Fault Tolerant Design of Multimedia Servers. In *Proc. of the ACM SIGMOD Conf. on Management of Data*, pages 364–375, San Jose, CA, May 1995.

[3] L. Golubchik, J. C.-S. Lui, and R. R. Muntz. Adaptive Piggybacking: a Novel Technique for Data Sharing in Video-on-Demand Storage Servers. *ACM Multimedia Systems Journal*, 4(3):140–155, 1996.

[4] S. W. Lau, J. C.-S. Lui, and L. Golubchik. Merging Video Streams in a Multimedia Storage Server: Complexity and Heuristics. *To appear in the ACM Multimedia Systems Journal*, 4(3):140–155, 1996.

[5] M. Y.-Y. Leung, J. C.-S. Lui, and L. Golubchik. Buffer and I/O Resource Pre-allocation for Implementing Batching and Buffering Techniques for Video-on-Demand Systems. In *Proceedings of the Intl. Conference on Data Engineering (ICDE '97)*, Birmingham, UK, April 1997.

[6] S. Marcus, V.S. Subrahmanian, and L. Golubchik. Sync Classes: A Framework for Optimal Scheduling of Requests in Multimedia Storage Servers. In *Proceedings of the 3nd Intl. Workshop on Multimedia Information Systems*, September 1997.

[7] M. Papadopouli and L. Golubchik. Support of VBR Video Streams Under Disk Bandwidth Limitations. *To appear in ACM SIGMETRICS Performance Evaluation Review*, 1997.