A Novel Approach for on-Demand Data Delivery System for Satellite Communications

John Blesswin
Karunya University
Tamil Nadu
India

Cyju Varghese
Karunya University
Tamil Nadu
India

Navitha
Karunya University
Tamil Nadu
India

Sonia
Karunya University
Tamil Nadu
India

ABSTRACT
Efficient load balancing is an important requirement for intelligent engineering of traffic over all-IP satellite networks. In a Satellite large-scale multimedia storage system (SLMSS) where client requests for different multimedia objects may have different demands, the placement and replication of the objects is an important factor, as it may result in an imbalance in server loading across the system. Since replication management and load balancing are all the more crucial issues in multimedia systems, in the literature, these problems are handled by centralized servers. Each object storage server (OSS) responds to the requests that come from the centralized servers independently and has no communication with other OSSs in the system. In this paper, we design a novel distributed load balancing strategy for SLMSS, in which OSSs can cooperate to achieve higher performance. Balance the requests among other servers to achieve a near-optimal average waiting time (AWT) of the requests in the system. Our proposed strategy is scalable, flexible, and efficient for real-life applications.

General Terms
On demand data delivery, Communication systems, Satellite communications, Replications

Keywords
Request balancing, Load balancing, Object storage servers, Replication, Multimedia System

1. INTRODUCTION
The dramatic change brought about with the advent of digital television is the conversion from an analog to a digital representation of the television spectrum. In the new digital world, data is easily and efficiently passed through from its source to destination in an unaltered fashion. As a result, data casting is rapidly achieving parity with video and audio services in digital television systems.

Sources of data are expanding (Internet, streaming video and audio services, digital archives) along with the public's demand for more and varied information-based products. The wide and easily accessible "data pipe" provided by digital television has placed data on a similar level of importance with video and audio services to the home. As a result, data casting has become the newest source of revenue-producing services for digital television programmers and distributors.

This paper explores the distribution and delivery of Internet Protocol (IP) data services, and the role of digital satellite and digital cable systems in the delivery of these services from the programmer to the viewer. Everyday lots of user clients are requesting for multimedia data form satellite server. A satellite large scale multimedia system (SLMSS), which consists of a single server or multiple servers, manages the storage and retrieval of multimedia data from disk media. As a networked multimedia service is expected to serve a large pool of clients, it is impossible for a single server to meet all multimedia requirements [5] such as continuous-time presentation network bandwidth, etc. Multi-gigabyte multimedia files are common, and the data sets are fast growing with daily added multimedia files of news, TV serials, etc.

System administrators have to regularly handle many terabytes of files. Due to their reliability and flexibility, multi-server storage systems have gained a predominant position in both the academe and industry. An alternative comes from the same programmers that already provide, via satellite, the vast majority of programming carried over cable systems today. Satellite delivery of IP data can bypass the router bottlenecks in the terrestrial Internet and deliver data directly to cable systems with a very high quality-of-service.

The programmers could provide program related IP services to augment the viewing experience of their existing programs, or use data to offer an expanded range of programming options to properly equipped cable subscribers. This would allow the cable operator to further differentiate his broadband services from those of the narrowband Internet, and be a further incentive for his subscribers to upgrade their set top terminals and be able to participate in more premium services. Servers can generally be classified in two large categories: centralized and distributed.

In a centralized system, a high-end scheduler monitors, serves, and manages the video streams among the servers whereas in a distributed system, [1] workstations or PCs can work as servers, and they cooperate with limited local information to achieve high system performance. Satellite delivery of real time video, audio and data is supported by Motorola's MPEG-2 encoder systems. Each encoder system provides up to eight MPEG-2 television channel encoders in a single chassis, with multiplexers, encryption, modulators and redundancy. Each MPEG-2 encoder channel includes a data channel suitable for transmitting synchronous data up to 9 Mbps with a constant delay through the satellite link.

This is referred to as isochronous data transmission. The high bit rate and constant delay provide the high quality of service desired for streaming data or large file downloads. As highlighted in the literature, compared with the centralized strategies, the distributed management of the system offers more advantages and, hence, more and more research works have been concentrating on distributed environments. For the traditional distributed computer networks, [4] each server in the system can provide computation capability for every client, whereas in multimedia storage systems, only the server, which has stored the requested objects in its disk(s), can provide the data retrieval service for the clients. It may happen that, assuming that a client request arrives at a server for a movie object, whose bandwidth has been allocated to other served movie streams, although there may be aggregate bandwidth available in other servers without the requested object, the request has to be blocked. One solution to solve such problems is to make some replications among other servers in the system so that other servers can serve the request. Hence, the placement and replication of the multimedia objects are important issues in a multimedia storage system.

As a networked multimedia service is expected to serve a large pool of clients, it is impossible for a single server to meet all multimedia requirements such as continuous-time presentation, network bandwidth, etc. In a large-scale multimedia storage system where client requests for different multimedia objects may have different demands, [5] the placement and replication of the objects is an important factor, as it may result in an imbalance in server [7] ing across the system. Generally each object storage server responds to the requests that come from the centralized servers independently and has no communication with other OSSs in the system. But in this distributed load balancing strategy for SLMSS, in which OSSs can cooperate to achieve higher performance [7]. Such OSS system can replicate the objects to and balance the requests among other servers to achieve a near-optimal average waiting time (AWT) of the requests in the system.

2. SCOPE OF THIS WORK

Depending upon the type of flows and the distribution of users across spot-beam coverage areas, the load on every spot-beam queue could be different. In a typical system, spot-beam queues share the access to antennas, transmitting in bursts when the connection is realized. If all spot-beam queues were given equal time shares, the amount of data that could be transmitted by a flow would be dictated by its minimum service rate among the queues which it belongs to. A high load variation among queues would under-utilize those queues that could serve the flow at a higher rate. We focus on dealing mainly with two conflicting issues that are of utmost importance in the context of satellite large-scale multimedia storage related problems.

Thus, first, our problem context considers a distributed system scenario. Although there exist a variety of performance metrics for evaluating such distributed strategies, based on the current context of the problem, a natural and primary choice is to consider three important metrics, namely, AWT, storage server usage/utilization, and Denial of Service (DoS). We will also include several different request arrival patterns within the scope of our simulation studies and attempt at putting forth certain key observations that are crucial to the design of such systems.

3. SYSTEM MODEL

In the following, we first present system architecture in the design of our strategy. Initially, some objects are assigned to a server, and this server becomes their primary server. All of the requests for these objects come to the primary server first, and the server takes charge of replicating the objects and distributing the requests among the system. Although it is a distributed system, due to the fact that the load balancing solution can be computed offline [6], each server can obtain some system information and can cooperate to achieve a global optimal solution. [2]

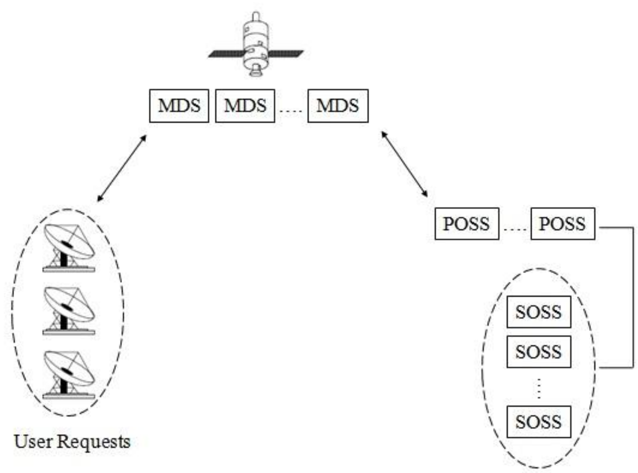

Figure 1. System architecture of our proposed SLMSS

We consider the system architecture of SLMSS as shown in Fig. 1, where we only present the main path of the requests and omit the technical details such Authentication, Authorization, and Accounting (AAA), the setup of secure sessions, etc. In this figure, there are three main steps for the requests: 1) the registered users send requests to the MDSs of SLMSS, 2) the MDSs forward the requests to the primary OSSs, which are in charge of the requested objects, and 3) one of the OSSs pushes the data stream to the client who has sent the request.

4. PROPOSED STATERGY

Two recent techniques for multicast or broadcast delivery of streaming media can provide immediate service to each client request, yet achieve considerable client stream sharing which leads to significant server, network bandwidth savings. This paper considers (1) how well these recently proposed techniques perform relative to each other, and (2) whether there are new practical delivery techniques that can achieve better bandwidth savings than the previous techniques over a wide range of client request rates. In SLMSS, if some servers remain idle, whereas others are extremely busy, the system performance will drastically be affected. We focus on popular, widely shared files, such as popular news clips, product advertisements, medical or recreational information, television shows, or successful distance education content, to name a few examples. [3] Due to the large size and the typical skews in file popularity, for the most popular

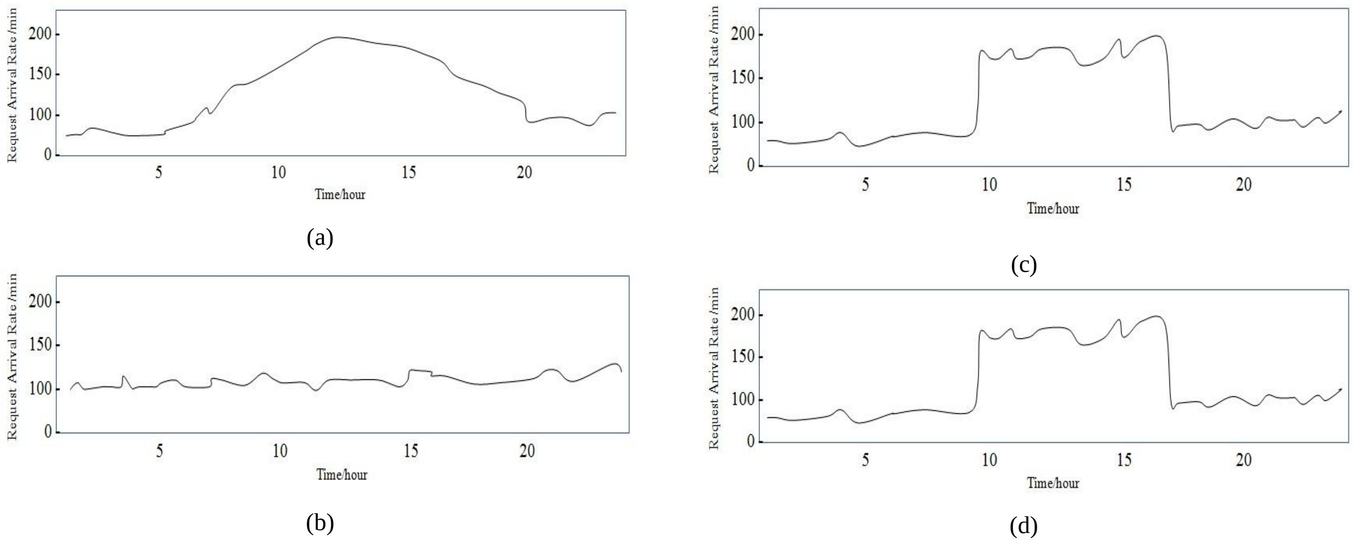

Figure 2. Various request arrival patterns. (a) Steady. (b) Gradual increase and decrease. (c) Square form. (d) Square form with hourly spikes.

files, one can expect many new requests for the file to arrive during the time it takes to stream the data to a given client.

To prevent this, load balancing is often used to distribute the requests and improve performance measures such as the average waiting time (AWT), which is the time difference between the time instant at which a request arrives to the system and the time instant at which the client starts receiving the first package of the multimedia stream In our proposed strategy, we distribute such workloads of load balancing among all the OSSs in the system and let the MDSs focus on searching for the locations of the requested objects only. Each OSS can determine the replications of the objects and balance the requests for these objects among their replications independently according to the near optimal solution of the objective function.

5. REQUEST ARRIVAL PATTERNS

As the Internet-connected population increases, we can observe that events external to the Internet conspire to create periods of significant fluctuation, and we shall employ a model for creating varied request arrival patterns. As in it is wiser to model the request arrival based on a 24-hour period. In Fig. 2, we propose a baseline steady-state request arrival pattern and three non-steady state request arrival patterns. Fig. 2a shows a steady state request arrival pattern, which is used as a comparison with the other patterns. For the non-steady state patterns, we use the statistics that suggests that the request arrival rates during the peak hour can approximately be four to five times greater than the normal rate. Figs. 2b and 2c show two patterns: one with a gradual increase to high rate and then a gradual decrease to low rate and one with a "jump" to a high rate. The pattern shown in Fig. 2d is a special case, with the rate surging at hourly intervals. Changes in the request arrival patterns are implemented using an exponential function with changing averages.

6. EXPERIMENT RESULTS

In this paper, we employ a discrete-event approach to model, simulate, and evaluate the systems. The multimedia servers are connected by an intranet or Internet. That the queue discipline of the requests in each server is FCFS. That the multimedia servers can be normal PCs or workstations, and the system can be homogeneous or heterogeneous. Each object is assigned to a server by a simple hashing function using the object ID as the key, and the server will become the primary server of the object. In order to guarantee the real-time delivery requirement of the continuous media, the servers have to provide each request with a minimum bandwidth for playback. That the delay of each communication such as polling sending, responding, and request forwarding follows a uniform distribution within [1, 6] ms, and the probability that the communication may be failed is set to 1 percent. We use synthetic request arrival patterns in our simulations on the performance comparisons. Two categories of request arrival patterns are generated for this purpose: 1) Steady requests, which include the steady arrival patterns, as shown in Fig. 2a, and 2) Dynamic requests, which include gradual increase and decrease arrival patterns, sudden increase and decrease arrival patterns, and sudden increase with hourly spikes arrival patterns, as shown in Figs. 2b, 2c, and 2d, respectively.

7. CONCLUSION

In this paper, a novel load balancing strategy has been proposed for Satellite large-scale multimedia storage systems, (SLMSS) which consist of a large number of OSSs. In traditional distributed file systems, request balancing is handled by the centralized MDSs, and the MDSs are the potential bottlenecks in achieving higher performance. In our proposed strategy, we distribute such workloads of load balancing among all the OSSs in the system and let the MDSs focus on searching for the locations of the requested objects only.

Each Object Storage Servers (OSS) can determine the replications of the objects and balance the requests for these objects among their replications independently according to the near-optimal solution of the objective function. For each object, it has a primary Object Storage Servers (POSS) in the system, which takes charge of the object. The primary server uses a heuristic method for finding other OSSs to store a replication of the object and construct the set of servers for the object.

The number of replications of each object is based on the request rate for the object. For example, some "hot" objects can have several replications to meet the client requests, whereas some "cold" object may have no replication, and they just have one copy for backup in the system. To the best of our literature survey, this work is the first of its kind, in which the request-based replication policy is proposed. This replication policy can use the disk storage capacity more efficiently and can save more disk space in the system.

8. ACKNOWLEDGMENTS

The Authors would like to thank the Innovative Transtar Research Team in Karunya University, for supporting us in our research work.

9. REFERENCES

[1] B. Veeravalli and G. Barlas, Distributed Multimedia Retrieval Strategies for Large Scale Networked Systems. Springer-Verlag, 2005.

[2] C.Y. Choi and M. Hamdi, "A Scalable Video-on-Demand System Using Multi-Batch Buffering Techniques," IEEE Trans. Broadcasting, vol. 49, no. 2, pp. 178-191, June 2003.

[3] W.K. Park, C.S. Choi, D.Y. Kim, Y.K. Jeong, and K.R. Park, "IPTVAware Multi-Service Home Gateway Based on FTTH Access Network," Proc. Ninth Int'l Symp. Consumer Electronics (ISCE '05), pp. 285-290, June 2005.

[4] U. Jennehag and T. Zhang, "Increasing Bandwidth Utilization in Next Generation IPTV Networks," Proc. Int'l Conf. Image Processing (ICIP '04), Oct. 2004.

[5] Y.J. Oyang, C.H. Wen, C.Y. Cheng, M.H. Lee, and J.T. Li, "A Multimedia Storage System for On-Demand Playback," IEEE Trans. Consumer Electronics, vol. 41, no. 1, pp. 53-64, Feb. 1995.

[6] B. Ozden, R. Rastogi, and A. Silberschatz, "Buffer Replacement Algorithms for Multimedia Storage Systems," Proc. Third IEEE Int'l Conf. Multimedia Computing and Systems (ICMCS '96), pp. 172-189, June 1996.

[7] Yun Sik Kim; Young-Ho Bae; Youngjae Kim; Chul Hye Park; "Traffic load balancing in low Earth orbit satellite networks" Computer Communications and Networks, 1998. Proceedings. 7th International Conference on Digital Object Identifier: 10.1109/ICCCN.1998.998776 Publication Year: 1998, Page(s): 191 – 195.