The ALMADEM Project
Applications and Analysis of Multimedia High Performance Networks

(Aplicações e Análise de Algoritmos de Redes Multimídia de Alto Desempenho)

Edmundo de Souza e Silva (coordenador) Sergio V.A. Campos
Rosângela Coelho Nelson Fonseca Rosa M. M. Leão
Berthier Ribeiro Neto Paulo Henrique A. Rodrigues
José Augusto Suruagy Monteiro Jorge Moreira de Souza
Sibelius Lellis Vieira
Wellington Santos Martins

UFRJ, CPqD, PUC-RJ, UFG, UFMG, UFPE, UNICAMP

Abstract

The field of Computer Networks has been growing at an enormous pace in the last several years. New technologies have been developed to take into account the stringent requirements of multimedia applications including audio and image, distance learning, collaboration and entertainment services. The main characteristics of multimedia applications are that they have vastly different performance and reliability requirements, often coupled with real-time constraints. Thus the design and evaluation of such multimedia systems and networks is quite a challenging enterprise. This paper surveys the main research achievements of the ALMADEM project financed by the ProTem-CC/CNPq program. The project includes both fundamental research in networks as well as the prototyping of designing tools and of multimedia applications. The participants are researchers from seven Brazilian institutions and foreign researchers from the USA and Europe.

1 Introduction

Computer Networks has been an area rich in successful initiatives and great social impact. The technological advances since the dawning of the ARPA-net, more than two and a half decades ago, where enormous and opened up the possibility of transmitting a variety of traffic from many different applications with drastically different requirements. Virtually all traditional concepts of network design have been re-examined in the last few years. The new, highly interactive, multimedia applications with vastly different performance and reliability requirements, often coupled with real-time constraints, led to a whole set of new problems that required new solution approaches different from those used in traditional networks. The design and evaluation of networked multimedia information systems is still filled with exceptionally challenging problems.

The main goal of the ALMADEM research proposal was to combine basic research in computer networks with the development of designing tools and laboratory tests to understand many networking design issues and propose new solutions. Yet we propose to build multimedia applications to test several of our ideas. In summary, the research proposal focused on a whole cycle of technology development, from theory to prototyping. The proposal can be divided into three major sets of activities:

1. *Fundamental Research in Computer Networks*, including: (a) development of new modeling and analysis techniques aiming at the evaluation of computer networks; (b) evaluation and development of basic control algorithms for multimedia networks; (c) traffic modeling.

2. *Modeling and analysis tools* that incorporate the state of the art in analysis techniques as well as new ones developed during the project. The tools would include analytical and simulation methods to be used in the analysis of high performance networks. Libraries of models and traffic simulators would be included. Testbeds would be designed to study the behavior of control policies used in high performance networks.

3. *Development of applications* including: (a) multimedia applications tools and network design tools; (b) a low cost Video on Demand (VoD) prototype.

Figure 1 shows the main tasks executed during the project. (This figure was taken from the proposal, and slightly modified to adapt to the changes of CPqD plans that occurred during the project.) The first group of tasks called *Network Fundamentals* is concerned with basic research in the area to enhance our fundamental understanding of networks. Key research topics that provide the foundation for transferring multimedia data over high performance networks were investigated. In this group, new modeling techniques were developed in order to allow more accurate network models to be solved. There are five major tasks in this set: (a) new analytical solution techniques and simulation methods; (b) traffic modeling; (c) models for network design; (d) specification of a VoD prototype.

The second group is called *Modeling tools*. The tasks in this group are highly coupled with those from the *Network Fundamentals* group and concerned with the design of tools that incorporate modeling and analysis techniques. Four major tasks are included in this set: (a) design of modeling tools; (b) design of model libraries used in the analysis of high performance networks; (c) design of traffic model libraries to be used in laboratory tests and simulations; (d) development of test-beds to collect data and test new ideas. The testbeds required the set up of laboratories in some of the institutions involved.

Finally, the third group of tasks called *Development of continuous media applications* is concerned with the prototyping of multimedia applications, thus bridging fundamental theory and network applications. This group includes video conference applications and the prototype of a VoD server.

The purpose of this paper is to briefly survey the main achievements of the project. Here we reference only our own work to emphasize our contributions. Details of the tasks executed, including the mapping of each task with the original proposal, the utilization of funds, papers published, difficulties encountered, change of plans with justifications, research groups that were created, cooperation between groups, students involved and international cooperation can be found in the final project report.

Section 2 describes the major results achieved in the first group of tasks mentioned above, including models and algorithms we developed. In section 3, we describe the tools we developed, and their main features. Section 4 presents two continuous media applications prototypes. Initially the design of the VoD prototype is described, including the main issues encountered. Then we briefly survey the unique features of the voice transmission tool developed. In section 5 we indicate in a few words, the role of the laboratories we built with the resources received. Finally, in section 6, we present our main conclusions, including the impact of the project in the groups, the projects that spawned from the ALMADEM project, and future research.

2 Network Fundamentals

The area of modeling and analysis has provided the theoretical basis for understanding the behavior of networks and for developing new algorithms and techniques to achieve the necessary network performance. The development of analytical performance models of complex systems is quite a delicate and creative undertaking. This requires a very careful balance between mathematical complexity and detailed description of the characteristics of the modeled system. Our contributions in this area can be divided into the development of network models, the development of new solution techniques to solve complex models of multimedia networks and traffic modeling.

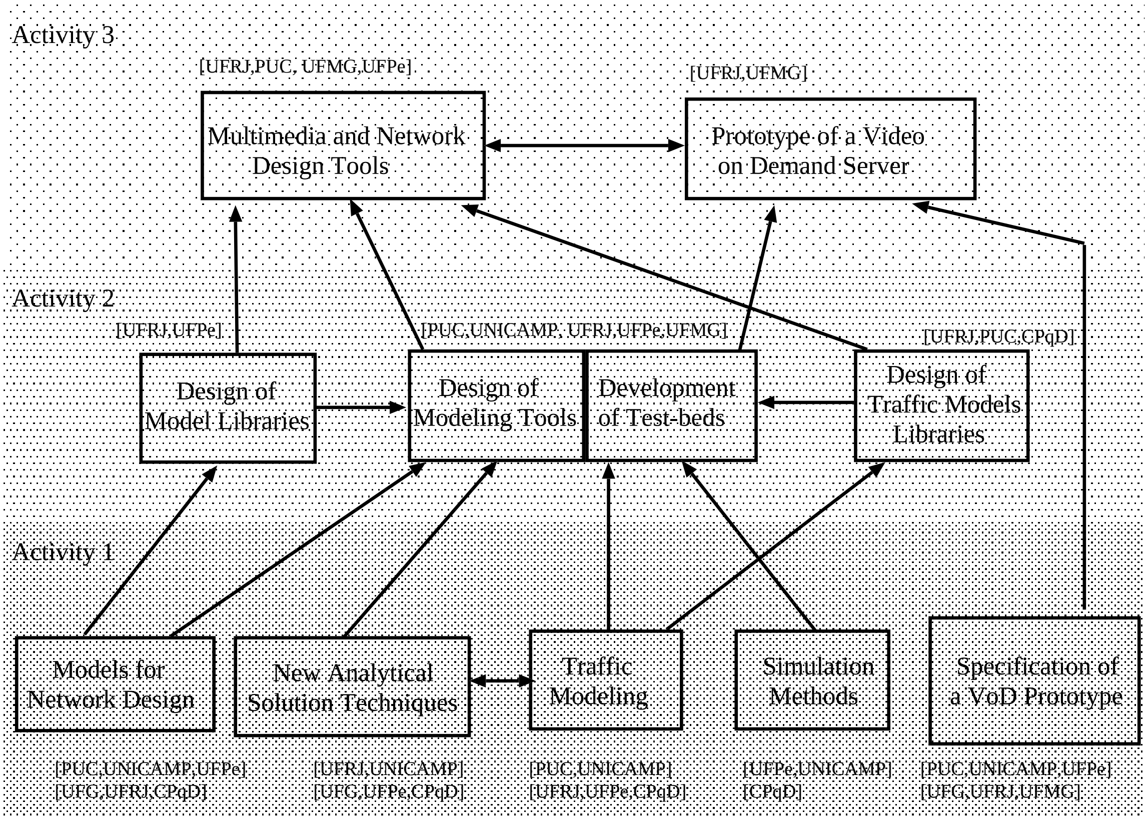

Figure 1: Main tasks of the ALMADEM project

2.1 Solution Techniques for Network Modeling

The area of computer networks is rich with examples that use basic as well as sophisticated modeling techniques to support new proposals, from protocols to control algorithms, that later became an industry standard. For instance, performance analysis has had a key role in the standardization of ATM technology. In the ATM forum, new proposals must be supported by performance studies, based on comprehensive models.

One of the major tasks of the ALMADEM project is the development of tools to solve complex models of multimedia networks. The tools would employ state of the art solution techniques to cope with the increasing complexity of models. In this section we summarize our contributions in the area of solution techniques.

Markov models which had been used earlier for a variety of applications, continue to play a major role for the performance evaluation of communication systems. We had previously developed a new and efficient solution techniques for Markovian models in which the resulting state transition matrix has special characteristics (e.g. block Hessenberg). These results are used in models of TCP over ATM [37]. These studies lead to new extensions and the results are included in [39].

We also investigated new approximate solution techniques for very large models based on the well-known GTH algorithm. Our findings show that one can reduce the cost of the solution by one on even 2 orders of magnitude for a class of network models, when compared to the GTH algorithm, and yet bound the error of the final solution. This research is part of the subject of a PhD dissertation that is in its final stages.

Equilibrium solutions are good approximations of the transient behavior of the system, provided that the convergence to steady state is fast in comparison to the observation interval. However, many examples exist for which it is imperative to perform transient analysis. Transient analysis [17] also plays an important role in the design and evaluation of control policies in communication networks. The need to design systems that degrade gracefully stimulated the definition of new metrics and the development of techniques which take into account, for instance, the combined effect of both performance and dependability (*Performability*). This class of metrics also result from fluid models, for instance when rewards representing packet arrival rates are associated with the states in the model. We developed a novel algorithm to calculate performability (transient) measures that includes both reward rates associated with states as well as impulse rewards associated with transitions [16]. These results were used in the solution of fluid models of multimedia networks [18].

We also developed a new approximation technique for the transient analysis of Markov reward models. Two solution techniques were considered: iterative an direct technique. Our studies show that the approach we developed is more efficient than traditional techniques specially when the model is *stiff* [21, 36].

2.2 Traffic Modeling

Traffic characterization is a capital issue for network performance analysis. For instance, traffic models are used to estimate buffer overflow probabilities, cell loss rate, cell discarding probabilities, and to evaluate connection admission control policies. Performance predictions of networks that carry multimedia traffic are relevant for maintaining the required quality of service.

A large number of models have been proposed to describe a wide variety of multimedia traffic. Ideally, these models should be general enough to represent the statistical behavior of a wide range of sequences (e.g., a video teleconference coded in H-261 or MPEG-1, an action movie picture coded in MPEG-2) and yet allow an accurate evaluation of the system performance parameters (e.g., cell loss probability).

In the ALMADEM project we proposed to investigate two promising techniques to model multimedia traffic: techniques based on Markov models and fractal models. This includes: (a) Markovian Models: Markov Modulated Poisson Process (MMPP), Video Markov Prediction Model (VMPM), ON-OFF, Birth-Death Process, Markovian-WEB, Pseudo-Markovian; (b) Non-Markovian: Fractional Brownian Motion (fBm) and Fractional Autoregressive (F-ARIMA). These techniques would provide a wide-area scenario for the performance analysis related to the main research topics earlier described. Other goals included the development of traffic simulators for source generation and model validation. The Fractional Brownian Motion model was implemented using an iterative method proposed by Mandelbrot.

Recent traffic measurements studies on packet networks have shown long–range dependence (LRD) characteristics. This low–frequency statistics should then be investigated to determine its impact on network performance. This dependence degree is represented by the Hurst parameter.

We have then evaluated the Hurst parameter from video sequences subjected to a rate control at the source. The results show that these sequences have negative dependence or high frequency characteristics. Several video sequences coded in H.261, MPEG-2 and CCIR Rec. 723 for TV/HDTV applications were studied [9, 12]. The statistical multiplexing effect from data models that exhibit long range dependency was considered in [14, 15]. The studies were important to further understand the impact of different models over network performance measures. The cell loss ratio analysis considering Markov and fractional Brownian Motion (fBm) models was investigated in [10, 11]. The results show that for small buffer size networks, the Hurst parameter has no effect on cell losses. Moreover, the Markov model better fitted the real video sequences statistics.

The precise estimation of end-to-end performances depends on the accurate representation of the flows in the network. Therefore, an algorithm was also proposed for modeling the output process of an ATM switching node, in which the priority level of the arriving cells are correlated [29, 30]. Moreover, we specified a framework for queueing networks with prioritized Markov modulated flows.

We have also obtained analytical expressions of traffic descriptors from multimedia fluid models. Previous work for important descriptors such as the IDC obtained expressions for a small class of models. As a result of our work general Markovian fluid models can now be considered [6, 20]. A detailed study aimed at investigating the influence of different traffic descriptors is current under way.

As a result of a joint study with CPqD, we studied a fluid traffic model for the telephone connections to Internet providers. The model captures long range dependency characteristics for the active periods of an Internet connection. The model is still being improved.

2.3 Network Modeling

The problems studied include: existing protocols performance over ATM networks, traffic control, bandwidth allocation and admission control, jitter control, multicasting. The set of problems investigated provided the foundation for the studies that were performed throughout the project.

Protocols developed for the Internet, such as TCP, will still remain in place for years to come and will continue to be used over heterogeneous networks including ATM networks. As we will see in future sections, our laboratories include a mixture of traditional as well as ATM routers. The issue is to study the main performance problems encountered by these protocols when used over such environments. In [38] we focused on the development of an analytical and a detailed simulation model for the performance analysis of TCP connections supporting long file transfers within high-speed ATM networks. Then, a detailed model validation was performed, comparing the analytical results with those produced by a very detailed simulation setup which included the officially distributed BSD 4.3 TCP-reno code. Numerical results indicate that the analytical model is able to capture the overall behavior of the system with respect to the performance measures considered (i.e., buffer size, throughput, window size). The cost is orders of magnitude less than the one necessary to obtain reliable estimates from the detailed simulation model. In [36] the transient analysis of this model was performed.

Controlling traffic remains a major issue in multimedia networks. The purpose of the studies performed in the project is to understand the main design issues in ATM networks. Different multimedia applications, such as voice and MPEG video transmission, have diverse quality of service (QoS) requirements. Coping with different requirements is a challenging task.

Selective discard is a congestion control mechanism aimed at enabling the network to deal with diverse loss requirements. In a selective discard mechanism cells are discarded in an overflow situation according to their priority level.

The effectiveness of selective discard mechanism in a ATM switch subject to a process with long-range dependencies is studied in [28] (see also [8]). Loss conserving disciplines are considered, and the impact of the input process traffic characteristics such as the Hurst parameter into the per class loss rate and the loss gap length is evaluated. We found out that the Complete Sharing discipline is clearly worth adopting whereas Complete Sharing with Guaranteed Queue Minimum may be not.

Selective discard mechanisms with multiple classes of traffic was considered in [32]. We analyzed the difference between multiple class selective discard under a short-range dependent and selective discard under a long-range dependent process.

A novel packet discarding policy which maximizes the cell good put, i.e., the ratio of outgoing good cells to the total number of cells that arrive at a queue was proposed in [33]. The impact of network load, mean packet size, packet distribution and the Hurst parameter into both the cell goodput and into the packet goodput was also evaluated.

Bandwidth allocation is intimately connected with the concept of effective bandwidth and with admission control with associated service requirements. The equivalent bandwidth of a connection (source) is a characterization of its demanded bandwidth. The QoS requirements are provided in the network based on statistical multiplexing .

In [41] an equivalent bandwidth was obtained for video sources from several coding mechanisms. Results for video sources using both Markovian and fractal models are presented in [9, 12].

In this project we also considered multiplexing policies with server timeouts for traffic with stringent delay and loss rate requirements. In [19] a few different policies are compared and we were able to analyze the efficacy of these policies over performance metrics such as delay and packet loss rate. This work requires the use of transient analysis, and the algorithm used was developed by us under this project.

The random variable defined as the interval between arrivals of cells minus the (constant) emission interval is called jitter and it is an important performance metric for real time voice applications. One application developed under this research is a real time voice tool and so we developed models to evaluate jitter control algorithms. At issue is whether the jitter control should be a network service or provided by the application. We advocate jitter control by the application based on previous results [25]. The models we developed show that a simple mechanism with introduces only an unperceptible delay on the voice packets can be used by the application. The analytical results obtained were compared with extensive measures made in the Internet between Brazil and the USA and other participant institutions of this project. The measured results supported our original claims [24, 23]. In [1] a different path was chosen and it was proposed a jitter reduction mechanism in the network based on service policies such as PGPS and Virtual Clock.

Another application developed under this project is a Whiteboard tool, as will be detailed below. One of the issues that needed to be addressed by this tool is the packet recovery mechanism for multicast transmission. In [7] the efficacy of different packet multicast retransmission algorithms is evaluated. It is important to emphasize that the model used in that work was developed and solved using the prototype modeling environment TANGRAM-II.

This modeling environment will be described in section 3.

2.4 VoD Modeling

One of the major applications developed in this project is a low cost VoD server. In [13, 40] a comparative study of several VoD systems architecture was performed. Since video applications it is likely to be one of the major bandwidth consumers of the future, it is also important to study mechanisms for reducing the bandwidth demand of video services. The studies would help to improve the usability of the prototype we developed. In a commercial VoD system, if one assigns a video channel for each individual random request, the number of active users is limited by the number of channels of a server. However, a large number of users can be supported in networks with multicast capability, by providing a single video stream to multiple users. Some mechanisms based on multicast such as piggybacking and batching have recently been proposed.

There are several contributions made in this area. A novel batching policy was proposed, called Look-Ahead-Maximize-Batch (LAMB), which maximizes the number of admitted users in a certain time window. We also analyzed the advantage of adopting batching instead of adopting piggybacking to considerably reduce the bandwidth requirements of VoD services. Moreover, we investigated the integrated use of batching and piggybacking.

Besides batching and piggybacking, one common technique to reduce the bandwidth demand of video service in large scale is to replicate servers along the network. We developed a framework to compute the network cost of video server replications which take into consideration the trade-off between replicating servers versus reducing bandwidth. We applied this framework to analyze the deployment of a video service with applications to distance-learning as well as professional training (see [26, 31, 34, 35]).

Finally, as it will be described in section 4.1, we were able to analyze the VoD prototype built. This analysis pointed out to important design directions.

3 Modeling Tools

This section describes in part the work associated with the rectangles in the central layer and part of the rectangles of the upper layer depicted in Figure 1. The tools we developed contains very sophisticated state of the art solution techniques and are useful to specify and solve models described in a high level language. We also developed tools that are integrated with multimedia applications.

3.1 TANGRAM-II

The TANGRAM-II modeling environment is based on an object oriented paradigm and it is developed for research and educational purposes [4, 5] [1]. It combines a sophisticated user interface and new solution techniques for performance and availability analysis.

In this paradigm, the system being modeled is represented by a collection of objects which interacts by sending and receiving messages. We choose to represent the internal state of each object by a set of integer valued variables. The behavior of an object is defined by events and messages and their associated conditions and actions. Events are generated spontaneously by an object, provided that the conditions specified when the object was defined are satisfied. These conditions are expressions evaluated using the current state of the object. Each event has a time-to-occurrence associated with it. Messages are just an abstraction used to represent the interaction among objects, and are delivered (and reacted to) in zero time.

The tool provides a graphical interface to aid the user when creating a new object type or defining a model. It is based on the public domain TANGRAM Graphic Interface Facility TGIF developed at UCLA. In order to create a new model, the user may employ different types of objects that can be retrieved from a library previously created, or he may build new object types from scratch. The objects are parameterized and instances can be declared with actual parameters specified for each instance.

Figure 2 presents the main modules of the tool architecture. The user starts building the model of a system by instantiating objects from the *Object Library* using the *Model Specification* module, provided that all type of objects needed are already in the library. Otherwise new object types are first created and stored in the *Object Library*. Then, the objects are parsed and data structures are created for each object instance. New objects types are defined as needed using the *Object Type Definition* module. In this module, new types of objects are specified in terms of a graphical representation and their associated behavior. The tool's interface displays a template to aid the user in the construction of a new object type.

After the model is built, the *Mathematical Model Generator* module generates a Markov chain, if the model is Markovian, and other information relevant to selecting a solution procedure. For non-Markovian models other data structures are generated. The user chooses the appropriate solution technique from the *Solvers* module to obtain the steady state and/or transient measures of interest. The options currently available in this module are shown in Figures 3 and 4.

[1] The paper [4] was chosen one of the best papers of the conference.

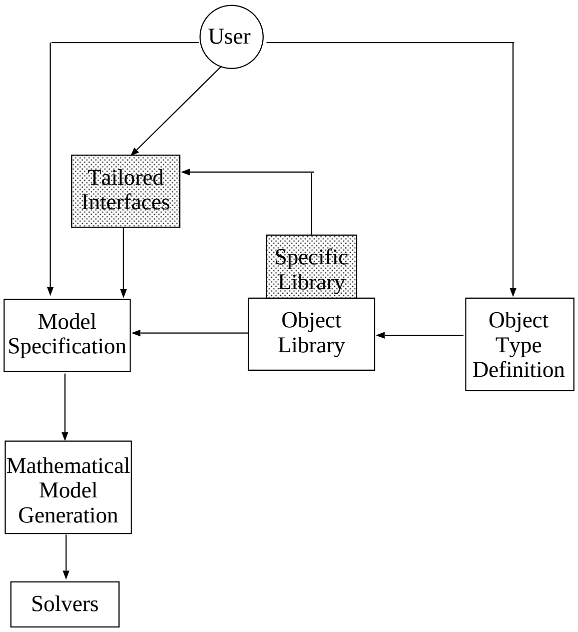

Figure 2: Tool Architecture.

During the scope of this project two versions of TANGRAM-II were generated. The first version included a simple parser and could only solve models with limited size, around a few thousand states. The tool was completely re-written to try to solve many of the problems encountered in the first version. Current features of the second version of TANGRAM-II include: a more sophisticated parser that allow object description using a C-like language; generation of large models (models with more than a million states); ability to generate symbolic transition matrices; specification and solution of reward models; new solvers; new interface to calculate different measures of interest. Figure 5 shows an example of a queueing model. In this example, an MMPP source feeds a leaky bucket traffic shaper and the resulting traffic is sent to a finite buffer queue. The input parameters are shown in the figure but the

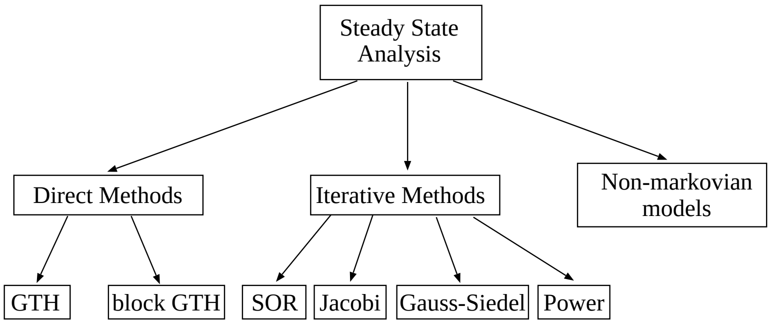

Figure 3: Module Solvers - Steady State Analysis.

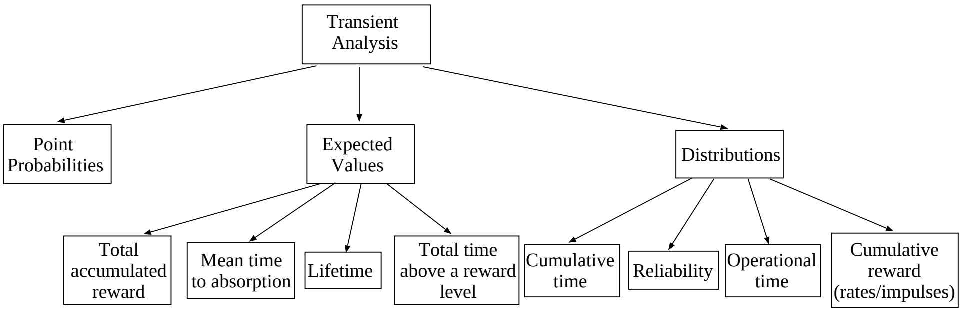

Figure 4: Module Solvers - Transient Analysis.

model description is hidden for conciseness.

A general purpose discrete event simulation was developed using the same language. This effort was carried out by requests from members of our group (CPqD, PUC-RJ and UFMG) such that models not solvable by analytical techniques could be solved by simulation in a integrated environment using the same interface. Our simulator also allows animation and step-by-step simulation to aid in the model debugging. Techniques for simulating models with rare events, are being currently incorporated.

We also developed and integrated the Whiteboard application to TANGRAM-II. This will facilitate the teaching of modeling and analysis techniques. Moreover, it enables cooperative work by providing an environment were users in different location could cooperate in building a model.

3.2 Library of Traffic Models

To support studies of multimedia networks, we developed a set of traffic models to be used in our overall network models. The new version of TANGRAM-II includes features such as: specification of inter event distributions including Gaussian, Fractional Brownian Motion, F-ARIMA, and capability to include general state transition models in an object description.

In order to illustrate one important modeling capability, suppose that a video sequence is coded with MPEG-1. Using our tools box we can obtain different models for the sequence that better match different traffic descriptors. These models can be fed into an object of a TANGRAM-II model and analytical and/or simulation studies can be done. We also obtained performance measures such as the cell loss ratio (for instance, assuming that the traffic model obtained feeds a finite buffer queue).

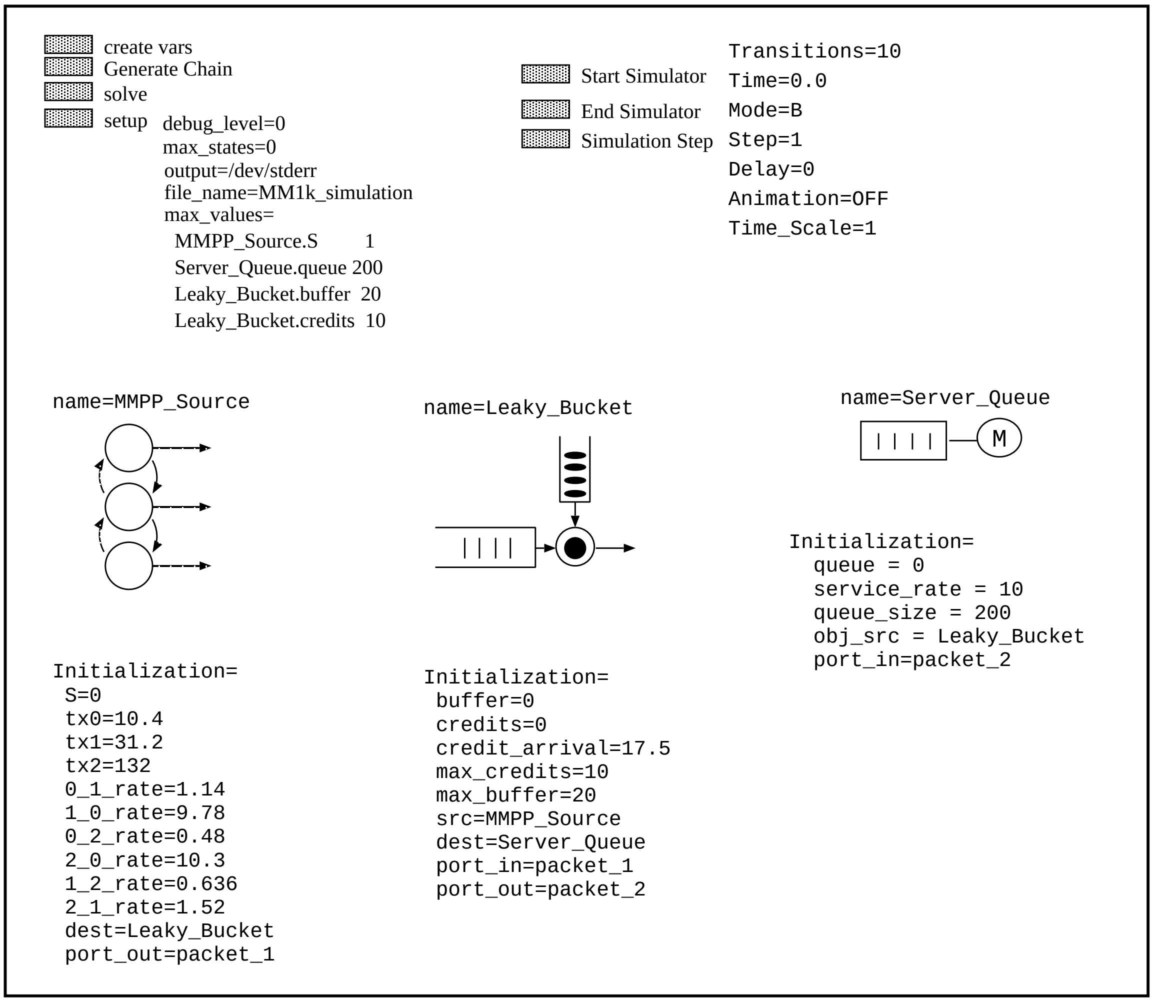

Figure 5: An example of a Queueing Model.

3.3 Whiteboard

As mentioned above the TANGRAM-II set of tools were developed to research and educational purposes. In order to further improve the usefulness of the tool box we develop a Whiteboard tool.

A Whiteboard tool, can be used in teleconferences and in cooperative work and is one of the applications with stringent QoS requirements. It requires the development of efficient recovering algorithms in an unreliable multicast environment and also ensuring data consistency between users. The implementation of a whiteboard would then also serve as a testbed to try many different algorithms for multimedia networks since the design issues involve reliable multicast transmission, a topic rich in research problems.

In [7] we describe the Whiteboard tool (Tangram Whiteboard - TGWB) developed in the scope of the ALMADEM project which has many features not available in other similar tools. The TGWB was implemented over a powerful public domain drawing tool. Its architecture is completely distributed, allowing simultaneous modifications in the drawing by any user in the group. Data consistency is insured in an unreliable multicast network.

We also developed an analytical model to study the sensitivity of performance measures to important design parameters for the packet recovery algorithm. The inconsistency problem of executing distributed graphic commands in different order is also address and a solution using total ordering of distributed events is developed.

3.4 Experimenta

The Experimenta (experimental design execution for performance evaluation analysis) tool was developed to facilitate the execution of performance modeling experiments, over different hosts on the Internet. Their results are depicted in statistical graphics and table form.

In more detail, Experimenta is a tool for automatically running batch jobs of a modeling design. It allows sweeps over arbitrary sets of parameter values. The run may be executed over the Internet on a set of hosts. The tool is written in Common Gateway Interface (CGI) scripts in the Perl language, returning dynamic documents written in JavaScript and HyperText Markup Language (HTML), and runs under Unix. It was originally designed for use with simulation program execution, but the tool can also be used with any test program that reads parameters from the command line or from an input file. The tool acts like an usual WEB tool that read and write documents with user access restrictions.

Experimenta is not yet integrated with the other tools in the project, but we foresee no difficulties in doing this task in the near future since Experimenta can be used with programs that reads parameters from an input file. However, we have not address this issue in detail.

3.5 VERUS

Verus is an efficient tool for performing the formal verification of multimedia systems and can exhaustively check the state space of systems with more than 10^{30} states, in a few seconds. It was developed prior the start of the ALMADEM project, but was later enhanced to fit the project needs to analyze the ALMADEM-VoD server. Verus represents the system being verified as a *state-transition graph* and verifies properties about its behavior described in temporal logic. It also allows determining *quantitative information* about the system such as its reaction time to events and the number of times an event occurs in a given time interval. The information produced allows the user to check the temporal correctness of the model and also provides insight into the behavior of the system. Such type of insight can help the user identify inefficiencies and suggest optimizations to the design. Further, this analysis can be performed *before* the actual implementation, significantly reducing development costs.

Verus has already been applied to the verification of several large complex systems (which are in production or are components of current industrial products) such as an aircraft controller, a robotics controller, and a distributed heterogeneous real-time system [2]. Recently, we have used Verus to verify the ALMADEM-VoD video server, and were able to find a design flaw [3].

Verus is a tool that complements the features of TANGRAM-II and we have plans to join Verus and TANGRAM-II in a unified environment. Currently we are exploring both tools in our research.

4 Continuous Media Applications

The fast development of new technologies for high bandwidth networks, wireless communication, data compression, and high performance CPUs has made technically possible the deployment of sophisticated infrastructures, for supporting a variety of multimedia applications. This type of infrastructure opens up opportunities for exploring multimedia applications such as quality audio and video on demand (from home), virtual reality environments, digital libraries, and cooperative design. Due to the high interactivity of such applications, the convenience of the service is heavily dependent on the delays incurred in the high bandwidth network, on the delays incurred at the server and client machines, and on the performance at the multimedia server.

4.1 Video on Demand Prototype

In here, we are particularly concerned with the performance of the multimedia server, in particular, of video on demand (VoD) servers. The design of the VoD server must consider basically the following delays: the delay to retrieve a block of data from disk, the delay to store the retrieved block in a memory buffer, and the delay to transfer a block from the memory buffer to the client machine (across the network). Further, the blocks of data must be transferred within strict time constraints, otherwise the visualization at the client machines will present discontinuities and *hiccups*. Typically, in a system that uses encoding in MPEG, the server must transfer to each client machine roughly 30 frames per second.

For MPEG-1, this implies in sustaining a bit rate of about 1.5 Mbps. Thus, in the design of a VoD system, we are mostly concerned with meeting the time requirements of the VoD application while attending as many clients simultaneously as possible. This is the so called *VoD problem*.

In the context of the ALMADEM project, we approached the *VoD problem* from a practical perspective. Our task within the project was to design, implement, and test a low cost VoD server. This effort lead to an efficient and reliable implementation which is referred to as the ALMADEM-VoD server.

4.1.1 Software Architecture of the ALMADEM-VoD Video Server

The fundamental premise in the development of the ALMADEM-VoD video server is that it should use only off-the-shelf low cost components. As a result, the server is implemented on a PC-based platform running the Linux operating system. To fulfill the real time requirements of the video application, the operating system is adapted in specific points such as the disk access and process scheduling routines.

Figure 6 illustrates the software organization of the ALMADEM-VoD video server. Two

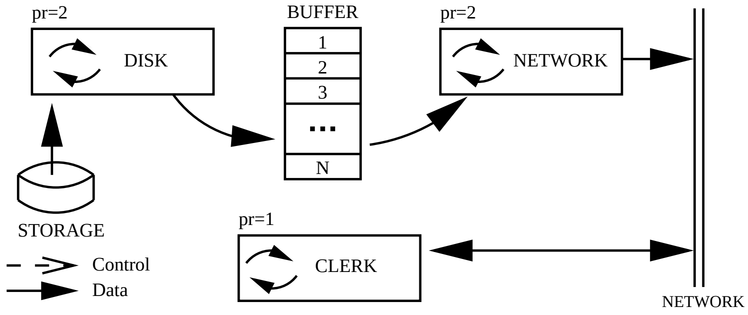

Figure 6: Software architecture of the ALMADEM-VoD video server.

data structures and three separate processes are distinguished. The data structures are called *storage* and *buffer*. The processes (implemented as POSIX *threads*) are called *disk*, *network*, and *clerk*.

The Linux scheduler has three different scheduling policies, named SCHED_FIFO, SCHED_RR and SCHED_OTHER. The SCHED_OTHER policy is related to the normal processes, those that do not have a hard urgency. The SCHED_FIFO and SCHED_RR policies are related to the "'real-time" tasks. Each task has an associated policy, a fixed priority and a variable priority. To ensure proper timing in the scheduling of the threads mentioned above, we use the SCHED_FIFO policy which implements a first-in-first-out scheduling scheme with static priorities. Despite its simplicity, this scheme works quite well if the machine is dedicated to the video server task (i.e., we run the video server in run time level 1).

The *storage* structure is composed of secondary or tertiary devices and is used to hold the collection of films available to the users. The current implementation of the ALMADEM-VoD server considers only secondary devices in the form of conventional SCSI-2 disks of 4G bytes each. The disks store the films encoded and compressed in MPEG-1 format. Each film is divided in *blocks* which are retrieved for delivery to the client machine. In its simplest implementation, which is adopted in this study, the ALMADEM-VoD server considers a *contiguous* layout of films in disk. In this layout, all blocks of a same film are stored contiguously on disk. More sophisticated layout schemes, involving striping techniques, region-based allocation, and randomized placement, have been discussed extensively in the literature but are not the focus of this work.

The *buffer* structure is basically main memory space used to synchronize the *disk* and *network* threads. It is implemented as a circular buffer which is filled by the *disk* thread and emptied by the *network* thread.

The *network* thread is responsible for taking the blocks of film from the buffer and shipping them across the network. It is scheduled whenever the *disk* thread is blocked at the disk driver waiting for a disk access to complete. This effect is accomplished by assigning the same priority to these two threads.

The thread named *clerk* listens at a TCP port for the requests from the client machines. Such requests might come from new clients or from a current client which requests, for instance, a *pause* in the exhibition. Once it detects a client request, this thread passes the information to an admission control routine for proper scheduling. If the server is saturated (i.e., it is currently serving a maximum number of clients), a request for a new stream is not scheduled and a denial message is sent to the respective client.

The *disk* thread is responsible for reading the data from the secondary storage and storing them at the buffer area. To avoid delays introduced by the operating system (which we cannot control), disk accesses are done through direct access functions which communicate directly with the SCSI controller device. For each active client, a separate block (which is composed of several MPEG frames) of (average) size B bytes is read, passed to the buffer area, and from there shipped (by the *network* thread) to the corresponding client machine. While that client consumes the frames in that block, other clients can be attended. This *cyclic scheduling* process is repeated with a fixed time period equal to T, as illustrated in Figure 7. To implement this fixed time period, the *disk* thread monitors the Real Time Clock (RTC) device in the Linux kernel. A seek time t_s to position the disk head at the proper cylinder and a transfer time t_r to read the block of data from disk (there is also a rotational delay which is not shown in Figure 7). The total time T_s spent serving all clients in the

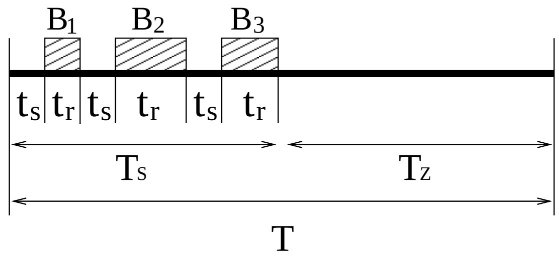

Figure 7: Service cycle with a duration of T seconds.

system is called the *service time*. The *sleeping time* T_z is the portion of the service cycle in which no clients are served. Such sleeping time is necessary because, to avoid buffer overflow at the client machine, the server does not attend a same client twice in a service cycle.

4.1.2 Experiments and Results

We verified and analyze the behavior of the ALMADEM-VoD server. We first investigated the occupation of the service cycle for different period values (T), with a fixed number of clients. For instance, we verified that the amortization of the seek time is minimal for periods larger than 7 seconds and does not improve much after 4 seconds. For periods larger than 4 seconds, we incur in higher latency without considerable additional gains in server performance. Thus, for the ALMADEM-VoD server, we adopt $T = 4s$ as a good compromise between client latency and server performance.

In Figure 8, we illustrate how the service time evolves as the number of clients in the system increases. The results are relative to operation of the ALMADEM-VoD server with a single SCSI-2 disk. The continuous curve is relative to measurements obtained from the ALMADEM-VoD server, while the dashed curves illustrate the minimum and maximum service times according to Verus, a tool for the formal verification of real time systems. Verus was used to verify the ALMADEM-VoD server. This verification provided a level of redundancy which was of great value during the debugging of the implementation. As observed, the service time for the ALMADEM-VoD increases linearly, as predicted by Verus. Further, it is within the minimum and maximum service times indicated by Verus. Since the period is 4000 miliseconds (i.e., $T = 4s$), we expect that the server will be able to attend a maximum of 50 clients on average. According to Verus, this maximum will be close to 35 clients in the worst possible situation and close to 70 in the best case scenario. More details can be found in [3].

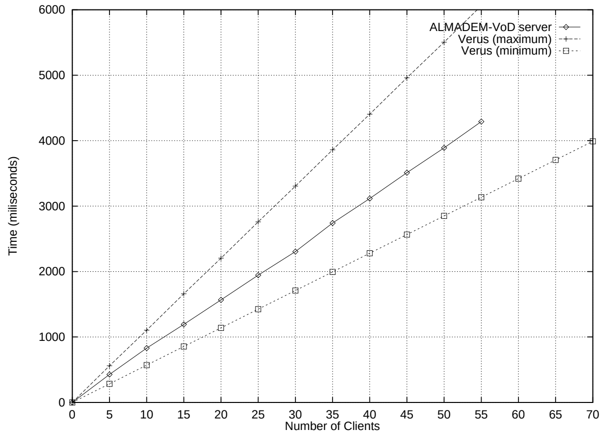

Figure 8: Variation of the service time with the number of clients for the ALMADEM-VoD server and for Verus. Period is $T = 4s$.

4.1.3 Remarks

The ALMADEM-VoD server design is in accordance with state of the art knowledge for the area. The implementation was done using only off-the-shelf components and the Linux operating system. The disk access and process scheduling routines of Linux are adapted to ensure that the time delays are within the bounds imposed by the video application.

We also report preliminary results on the behavior of the ALMADEM-VoD server. Our results indicate that the server, operating in a single standard PC machine, is able to support about 50 streams simultaneously when a single SCSI-2 disk is used. With two disks, the number of streams in the system should go up to 100.

We are also envisioning the use of the server to store different class of multimedia data with different class of requirements. Preliminary theoretical studies of scheduling policies for mixed workload can be found in [22].

The strategy of using low cost machines presents interesting advantages from a commercial point of view. First, we can experiment the service commercially with a small initial investment in hardware for the server itself. Second, we can increase the system capability incrementally such that the additional investment is proportional to the required improvement in performance. Third, in a wide area enterprise, we can distribute various small servers close to the premises of the users which reduces significatively the requirements in network bandwidth.

4.2 Voice transmission tool

Besides the VoD application described in the previous section we were able to develop another continuous media applications: a voice transmission (voice over IP) tool. This section briefly describes its main features.

Applications for transmitting voice over the Internet have been developed for some years, and today, a wide variety of these applications, commercial or freeware, are available. Some examples of freeware applications are the Vat tool, Nevot, Rat and FreePhone. The purpose of the voice tool developed was to provide us with the necessary "hooks" to perform controlled tests with real time applications and implement new ideas to cope with the problems encountered. It would also serve as a communication device among members of our group.

The unique features of the voice tool include a module to collect a variety of statistics from the network, a jitter recovering algorithm with parameters set by our analytical studies and a sophisticated packet recovering algorithm, developed by us to cope with heavy network losses common in our Internet environment of today. Details of the tool and the recovering algorithm used can be found in [27, 24, 23].

5 Developing Laboratories

The ALMADEM project aimed at establishing three laboratories at UFRJ, UFMG and UFPE. In these laboratories we would develop testbeds and a number of experiments to studies issues related to high-speed multimedia networks. It is interesting to observe that putting a state of the art network laboratory into full operation is not an easy task and requires expertize not readily available among our students. We were able to train a group of undergraduate and master students that now has the necessary ability to fully operate the labs.

Presently one FORE ASX-1000 ATM switch and one FORE ES3810/CH Ethernet switch in each lab is in operation. In the lab, PCs with Linux and FreeBSD and Solaris workstations are both ATM and Ethernet interconnected. We tested Lane, Classical IP and PVCs connectivity with success. VoD application operates directly over IP on these environments.

We will further deploy RSVP over ATM, on both Linux and Solaris environments, and use LANE 2.0 and/or MARS service to support Ipv4 or Ipv6 multicast with QoS. We plan to develop our own MARS implementation and use standard RSVP over ATM APIs in other to support direct QoS operation of our VoD application.

Furthermore, we have studied the integration of short-cuts provided by NHRP over ATM with RSVP operation for multicast RSVP sessions. Implementation of these new procedures will be a lab task for the future.

We have also the capability to tape movies and convert to MPEG in a reasonable time frame. The coded movies can be stored in our VoD server, and be used in our experimental studies. We can create a variety of traffic models and transmit the traffic generated by these models over Ethernet or ATM network. We are now in the position of perform a full range of analytical and experimental studies related to CAC admission optimization and ATM traffic modeling.

6 Conclusions and Future Work

In the ALMADEM project proposal we foresaw the execution of several tasks that ranged from the study and development of modeling techniques used in the analysis of high performance networks to the implementation of design tools and a VoD prototype. The vast majority of our initial goals were achieved. The only deviation from our plans resulted from the change in CPqD research interests, due to its privatization, while this project was being carried on. Even though, we were able to find new challenges of common interest, and the interaction resulted in theoretical technology transfer from the University to CPqD in the area of traffic modeling.

Our group was able to obtain novel results as can be seen by the list of international publications. Furthermore, we worked towards prototyping new design tools and continuous media applications that included several of our research findings.

Besides the theoretical results obtained that improved our understanding of high performance computer communication systems, we have prototyped important tools and applications. The TANGRAM-II environment has already been used in projects and/or undergraduate courses at UFRJ, UFPE and UFG. It has also been used in graduate courses at UFRJ and to obtain recent results for some of our papers. As soon as an english tutorial of TANGRAM-II is completed, the tool will be distributed worldwide. Many researchers have already expressed their interest in using TANGRAM-II. The voice and whiteboard tools will provide extra support for cooperative work and for distance learning.

The VoD server has demanded a major effort from our group and it is one of our most important deliverables. We were able to demonstrate the feasibility of design a low cost server with off-the-shelf components. The server will be continuously improved and in the mean time used in our labs for our tests and teaching courses.

One interesting result is that we were able to use one of our tools (VERUS) to analyse the VoD prototype and to indicate design directions to be followed. We also plan to develop a TANGRAM-II model in the future, as the features needed to obtain a detailed model are now implemented.

We have contacted the Institute of Thorax Diseases at UFRJ (IDT) and demonstrated the capabilities of the VoD server as a repository for telemedicine lectures that are currently being broadcast from the Mayo Clinic in the USA to UFRJ. A demo was made both to IDT and the Mayo Clinic during the visit of a Mayo Clinic delegation to UFRJ. As soon as the VoD prototype is put into full operation, it will store digitalized lectures from the Mayo Clinic to be accessible to doctors at UFRJ, initially, and doctors in other locations in the future.

We were able to build state of the art laboratories that have now the capability to perform many advanced research experiments. As a result of our theoretical studies, we have now the ability to generate a variety of traffic models from any stored movies digitilized in our labs. These traffic models can be fed into our network to provide the necessary environment for our studies. Graduate courses are currently benefiting from these laboratories.

Another important achievement obtained was the creation of research groups in the institutions involved. Some of these groups now have several undergraduate and graduate students involved in many different parts of the project. The cooperation between groups was not small either. In our final report we highlight these interactions including those among graduate students.

International cooperation was also pursued by different members in our group. Some of the senior students had benefited from this cooperation and were active in discussing research with professors and students abroad.

The ALMADEM project also spawned new research interests. We have compiled a list of new topics that include enhancements made to the set of modeling tools and to the VoD server. We will investigate the use of encoding schemes for transmitting video using very low bit rate over heterogeneous networks. Thus, enlarging the accessibility of the movies stored in our VoD server.

Since the ATM switches bought only arrived a few months ago, we have plans to perform many studies using the switches. We are current in touch with the switch manufactory to provide us with the necessary information for our tests. These tests will include studies with resource reservation protocols, integration of the ATM with digitalized telephone switch, study and implementation of different call admission control policies for the VoD server, implementation of shared streams for the VoD and integration of a real-time scheduler for the VoD. We are current writing a new research proposal to continue our work. Furthermore, a subset of the members in our project group have submitted a joint CNPq/NSF proposal in the area of continuous media applications. The proposal is currently being evaluated.

References

[1] L. Bermeo and R. Coelho. Análise de mecanismos de recuperação de variação estatística de retardo em tráfego multimídia para redes do tipo atm. *Anais do XV Simpósio Brasileiro de Telecomunicações*, 1, September 1997.

[2] S. Campos and O. Grumberg. Selective quantitative analysis and interval model checking: Verifying different facets of a system. In *Conference on Computer-Aided Verification*. Springer-Verlag, 1996.

[3] Sergio Campos, Berthier Ribeiro-Neto, Luciano Bertini, and Autran Macedo. Formal verification and analysis of multimedia systems. 1999. Submitted to Seventh ACM Multimedia Conference (Multimedia'99).

[4] R.M.L.R. Carmo, L.R. de Carvalho, E. de Souza e Silva, M.C. Diniz, and R.R. Muntz. TANGRAM-II: A performability modeling environment tool. In *Lecture Notes in Computer Science 1245: Proceedings of the 9th International Conference of Performance Evaluation Modelling Techniques and Tools*, volume 1245, pages 6–18, St. Malo, France, June 1997. Springer.

[5] R.M.L.R. Carmo, L.R. de Carvalho, E. de Souza e Silva, M.C. Diniz, and R.R. Muntz. Performance/Availability Modeling with the TANGRAM-II Modeling Environment. *Performance Evaluation*, (33):45–65, 1998.

[6] R.M.L.R. Carmo, E. de Souza e Silva, and S.C. de Lucena. Cálculo de Descritores de Tráfego em Modelos Markovianos de Fontes Multimídia. In *Anais do* 15º *Simpósio Brasileiro de Redes de Computadores*, pages 189–204, São Carlos, SP, May 1997.

[7] D.J. Oliveira e E. de Souza e Silva C.E.F. de Brito, R.S. de Moraes. Comunicação multicast confiável na implementação de uma ferramenta whiteboard. In 17º *Simpósio Brasileiro de Redes de Computadores*, Salvador, BA, Brasil, May 1999.

[8] R. Choen, N.L.S. Fonseca, and M. Zukerman. Traffic Management and Network Dimensioning. In M. Tatipamula and B. Kasnabish, editors, *Multimedia Communications Networks: Technologies and Services*, pages 133–173. Artech House, 1998.

[9] R. Coelho and L. Decreusefond. Fractional brownian motion usage for smoothed video traffic bandwidth allocation. *Proceedings of GLOBECOM'97*, November 1997.

[10] R. Coelho, L. Decreusefond, and S. Tohmé. Smoothed video traffic modeling for statistical multiplexing analysis. In *Proceedings of the 6th IFIP Workshop on Performance Modeling and Evaluation of ATM Networks*, Ilkley, UK, July 1998.

[11] R. Coelho, L. Decreusefond, and S. Tohmé. Statistical performance of tv/hdtv traffic over broadband digital networks. In *Proceedings of IEEE GLOBECOM98*, Sidney, Australia, November 1998.

[12] R. Coelho and S. Tohmé. A generic smoothing algorithm for real time variable bit rate video traffic. *Computer Networks and ISDN Systems*, 29:2053–2066, April 1998.

[13] Edilayne Meneses da Silva. Estudo comparativo de arquiteturas de sistemas de vídeo sob demanda. Master's thesis, Departamento de Informática, UFPE, February 1997.

[14] Adriano Lorena Inácio de Oliveira and José Augusto Suruagy Monteiro. Modelos de tráfego para a multiplexação estatística do tráfego de dados em redes ATM. In *Anais do 16. Simpósio Brasileiro de Redes de Computadores*, pages 703–722, Rio de Janeiro – RJ, May 1998.

[15] Adriano Lorena Inácio de Oliveira and José Augusto Suruagy Monteiro. Statistical multiplexing of aggregate data traffic in atm network. In *Proceedings of SBT/IEEE International Telecommunications Symposium (ITS'98)*, pages 108–113, São Paulo, August 1998.

[16] E. de Souza e Silva and H.R. Gail. An algorithm to calculate transient distributions of cumulative rate and impulse based reward. *Stochastic Models*, 14(3):509–536, 1998.

[17] E. de Souza e Silva and H.R. Gail. Transient Solutions for Markov Chains. In W. Grassmann, editor, *Advances in Computational Probability*. Kluwer, 1999. to appear.

[18] E. de Souza e Silva, H.R. Gail, and J.C. Guedes. Transient distributions of cumulative rate and impulse rewards with applications. In *Proceedings of the 18th IFIP TC7 Conference on Systems Modelling and Optimization*, Detroit, USA, July 1998. Addison Wesley Longman. 9 pages.

[19] E. de Souza e Silva, H.R. Gail, and R.R. Muntz. Gated time-limited polling systems. In *Proceedings of the International Conference on the Performance and Management of Complex Communication Networks*, pages 251–270, Japan, November 1997.

[20] E. de Souza e Silva, R.M.M. Leão, and S.C. de Lucena. Calculating Traffic Descriptors from Markovian Models of Multimedia Sources. In *Proceedings of INFORMS'98*, TelAviv, Israel, June 1998. (Extended Abstract).

[21] E. de Souza e Silva, R.M.M. Leão, and R. Marie. An efficient approximation technique for calculating transient reward measures. In *Proceedings of The Fourth International*

Workshop on Performability Modeling of Computer and Communication Systems (PM-CCS4), pages 16–19, Williamsburg, USA, September 1998.

[22] Edmundo de Souza e Silva, H. Richard Gail, Leana Golubchik, and John C. S. Lui. Analytical models for mixed workload multimedia storage servers. *Performance Evaluation 1999*. To appear.

[23] B.R. Brandi D.R. Figueiredo, E. de Souza e Silva. Efficient mechanisms for recovering voice packets in the internet. In *Globecom 99*, Rio de Janeiro, Brazil, December 1999. Submitted for publication.

[24] B.R. Brandi e E. de Souza e Silva D.R. Figueiredo. Mecanismos para recuperação de perdas de pacotes de voz na internet. In 16º *Simpósio Brasileiro de Redes de Computadores*, pages 52–67, Rio de Janeiro, Brasil, May 1998.

[25] M.C. Diniz e E. de Souza e Silva. Models for jitter control at destination. In *Proceedings of IEEE-ITS'96*, pages 118–122, Mexico, October 1996.

[26] R.A. Facanha, N.L.S. Fonseca, and P.J. Rezende. *Multimedia Tools and Applications*, chapter The S2 Piggybacking Policy. To appear.

[27] D.R. Figueiredo, F.P. Duarte, and E. de Souza e Silva. Implementação de um aplicativo de voz com análise e caracterização de seu tráfego. In *XXIV SEMISH, XVI CTIC*, pages 531–540, Brasília, DF, August 1997.

[28] N. L. Fonseca, M. J. Ferreira, and G. S. Mayor. On the Effectiveness of Loss -conserving Disciplines under a Long-range Dependent Process. In *Conferência IEEE Global Telecommunications Conference'97 (IEEE Globecom'97)*, pages 1102–1107, November 1997. Phoenix - Arizona.

[29] N. L. Fonseca and J. A. Silvester. Modelling the output process of an ATM mutiplexer with correlated priorities. *in Proc. of IEEE IEEE International Conference on Communications ICC'97*, pages 816–821, 1997.

[30] N. L. Fonseca and J. A. Silvester. Modelling the output process of an ATM mutiplexer with correlated priorities (extended version). *Computer Networks and ISDN Systems*, 29(26):2153–2164, February 1998.

[31] N. L. S. Fonseca and C. M.R. Franco. Using distributed server to provide distributed home theatre. In *Proceedings of IEEE International Telecommunications Symposium*, pages 261–266, 1998.

[32] N.L.S. Fonseca and M.J. Ferreira. Multiple class selective discard under a long-range dependent process. In *Anais da conferência IEEE International Telecommunications Symposium 98*, pages 95–101, 1998.

[33] N.L.S. Fonseca and M.J. Ferreira. On the effectiveness of the longest-packet-in packet discard policy. In *Anais da conferência IEEE Global Telecommunications Conference (IEEE GLOBECOM98)*, pages 1747–1753, 1998.

[34] N.L.S. Fonseca and C.M.R. Franco. Fornecendo serviços de distributed home theatre em redes com multicast e com servidores distribuídos. In *Anais do 16 Simpósio Brasileiro de Redes de Computadores*, 1998.

[35] N.L.S. Fonseca and C.M.R. Franco. A política de piggybacking s2. In *Anais do IV Simpósio Brasileiro de Sistemas Multimídia e Hipermídia*, pages 125 –136, 1998.

[36] R.M. Leão and J.A.N da Silva. Estudo de um método para análise transiente de modelos markovianos de sistemas de comunicação. In *Anais do* 16º *Simpósio Brasileiro de Redes de Computadores*, pages 742–761, Rio de Janeiro, RJ, May 1998.

[37] M. Ajmone Marsan, E. de Souza e Silva, R. Lo Cigno, and M. Meo. Approximate performance analysis of TCP over ATM. In *13th United Kingdom Workshop on Performance Engineering of Computer and Telecommunication Systems (UKPEW'97)*, pages 18/1–18/15, July 1997.

[38] M. Ajmone Marsan, E. de Souza e Silva, R. Lo Cigno, and M. Meo. A markovian model for tcp over atm. *Telecommunication Systems Journal*, 1999. to appear.

[39] M. Meo, E. de Souza e Silva, and M. Ajmone Marsan. Exact and Approximate Solutions for a Class of Infinite Markovian Models. Submitted to 3rd International Meeting on the Numerical Solution of Markov Chains.

[40] Carlos Marcelo Dias Pazos, Edilayne Meneses da Silva, Mario Gerla, José A. S. Monteiro, and Luigi Fratta. Video on demand distribution over ATM virtual private networks. In *Lecture Notes in Computer Science: Multimedia Applications, Services and Techniques — ECMAST '97*, volume 1242, pages 229–244, Milan, Italy, May 1997.

[41] L. Vieira and R. Coelho. Alocação antecipada de banda passante para fontes de vídeo vbr com qualidade de serviço garantida. *Anais do XV Simpósio Brasileiro de Telecomunicações*, 1, September 1997.