TECHNICAL RESEARCH REPORT

A Methodology for Modeling, Performance Analysis, and Control of ATM Networks with Multi-Media Traffic

by C-H. Chou, W-C. Chan, E. Geraniotis

CSHCN T.R. 97-20
(ISR T.R. 97-52)

The Center for Satellite and Hybrid Communication Networks is a NASA-sponsored Commercial Space Center also supported by the Department of Defense (DOD), industry, the State of Maryland, the University of Maryland and the Institute for Systems Research. This document is a technical report in the CSHCN series originating at the University of Maryland.

Web site http://www.isr.umd.edu/CSHCN/

A METHODOLOGY FOR MODELING, PERFORMANCE ANALYSIS, AND CONTROL OF ATM NETWORKS WITH MULTI-MEDIA TRAFFIC *

Chih-Hsien Chou
Wai-Chung Chan
Evaggelos Geraniotis

Department of Electrical Engineering
and Institute for Systems Research
University of Maryland, College Park, MD 20742

ABSTRACT

In this paper we review recent advances in developing a methodology for traffic modeling, performance evaluation, and control of ATM networks that can be used to support several aspects of Task 4.3 of the ATIRP project. Our methodology includes model matching and validation for multi-media traffic, analytical approximation techniques for time-efficient and accurate evaluation of end-to-end as well as intermediate node performance measures of multi-hop networks, and optimization of control (resource allocation) schemes.

INTRODUCTION

In the future battlefield for the Army, multi-media information should be distributed efficiently and reliably from the information sources to every desired destination in order to facilitate the tactical/strategic decisions-making process. This goal is most readily fulfilled by using modern digital communication technology which suggests digitize and encrypt the information before transmission. Among the multi-media information to be distributed are data, voice, image, and video. The current trend in modern telecommunication is that different kinds of these information will be segmented into fixed-length packets and transmitted via an integrated broadband network. Asynchronous transfer mode (ATM) and fast packet-switching are most suitable to be applied in this integrated broadband network because multi-media traffic with various characteristics and quality of service (QoS) requirements can be transmitted efficiently. In multi-media ATM networks, the availability of accurate performance measures or QoS is essential in optimizing the throughput of the network. As a consequence, accurate techniques to calculate the QoS of the network as a function of the network parameters and traffic characteristics are needed. However, their exact values are difficult to evaluate. This difficulty is due to the presence of different traffic types in the network, the large network size and its architecture, and the interdependence between different links in the network. Although exact expressions for the QoS assuming certain analytical traffic models are available in some cases, the computational efforts required are prohibitive even for small-size networks.

In this project, recently developed multi-media traffic modeling and analytical approximation techniques are used in order to develop accurate and time efficient approximations to the end-to-end performance measures (end-to-end QoS) of multi-media ATM networks. Two paradigms are considered in this project: (1) A cell-switched network with virtual path; (2) A virtual circuit-switched network with multicasting. For the first paradigm, end-to-end packet loss probabilities and packet delay times are calculated for each traffic stream. For the second paradigm, end-to-end call blocking probabilities are calculated for each traffic stream. Resource allocation schemes are also discuss for the first paradigm.

ATM NETWORK TRAFFIC MODELING

Recent research results show that most of the LAN traffic and VBR video traffic, which will be typical of the traffic streams present in the future ATM networks, are long-range dependent (LRD) in nature [1-3]. This leads to new insights into the problems of traffic modeling. We use four traces of actual traffic measurement

*Prepared through collaborative participation in the Advanced Telecommunications/Information Distribution Research Program (ATIRP) Consortium sponsored by the U.S. Army Research Laboratory under Cooperative Agreement DAAL01-96-2-0002.

data: (1) intra-frame encoded VBR video traffic, (2) MPEG1 encoded VBR video traffic, (3) Ethernet internal packet traffic, and (4) Ethernet external packet traffic, to access the statistical features of the traffic and estimate the parameters of the traffic models. Table 1 summarize some statistics of the four data traces. Different kinds of data processing methods are also used to change the characteristics of the actual data traces in order to test the relative importance among various aspects of traffic statistical properties on ATM network performance. The effect of traffic higher-order statistics on network performance is compared with that of second-order statistics and marginal distribution. The simulation results are presented in leaky-bucket contour (LBC) plots, which is an effective way to describe the characteristics of a bursty traffic source. In the simulation results we obtained [4], we can conclude that: (1) Gaussian process with the same second-order statistics as the empirical second-order statistics of the actual traffic data may not be a good traffic model as far as the network performance is concerned, because it does not model the marginal distribution well. (2) Higher-order statistics only have minor effects on the network performance compared with that of second-order statistics and marginal distribution. (3) Traffic energies in different portions of the traffic power spectral density have effects on the network performance in different resource allocation regions.

Trace	1	2	3	4
Data Length (sample)	171000	174136	175821	122797
Measure Rate (sample/sec)	24	24	100	1
Mean Rate (bytes/sample)	27791	1950	3630	1142
Peak/Mean Ratio	2.85	11.9	3.75	67.5
Hurst Parameter	0.78	0.90	0.78	0.86

Table 1: Statistics of the four data traces.

With these results in mind, we use heterogeneous on-off sources to synthesize simulated traffic which will match both the marginal distribution and the second-order statistics of the actual traffic data. Figure 2 shows the concept of an on-off source model. On-off sources with both exponential and Pareto distributed sojourn times are used and the resulting simulated traffic is the aggregated traffic of a number of heterogeneous on-off sources. Figure 3 shows the process of model parameter matching. Network performances for actual traffic data and simulated traffic data using different traffic models are obtained by simulation. From these results [5], we are able to identify the resource allocation region for each traffic model to be appropriately used. The validity of the proposed traffic models to characterize essential properties of the actual traffic can thus be confirmed. Methods for mapping the proposed heterogeneous on-off sources models to Markovian Modulated Poisson Process (MMPP) models are provided. The tractability of the analytical solving method using MMPP traffic models to obtain the relevant network performance is also explored.

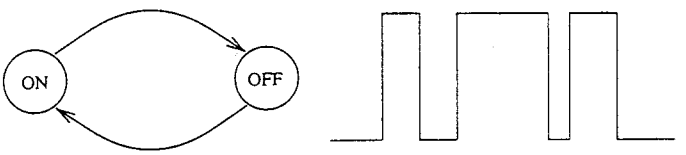

Figure 1: An on-off source model.

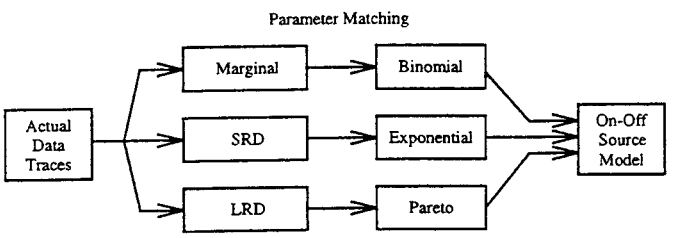

Figure 2: Traffic modeling using on-off sources.

PERFORMANCE ANALYSIS OF ATM NETWORKS AND SWITCHES

MMPP models and analytical approximation techniques are used in order to develop accurate and time efficient approximations to the end-to-end performance measures (QoS) of generic ATM switches and examples of ATM networks with multi-media traffic. Packet switched network with virtual paths is used as the network model. The algorithm used is based on the reduced-load method [6]. Packet dropping at different links are assumed to be independent initially and the dropping rates at the link-level are computed individually. These link-level dropping rates are then applied to compute the reduced load of the traffic. The process is iterated and will lead to a fixed-point problem. The solution corresponds to the link-level performance measures of each link in the network. The end-to-end packet dropping rates as well as other end-to-end performance measures of interest follow immediately. The

link-level packet dropping rates are computed using a simplified source model. The idea behind the simplified model is that by using state aggregation techniques, a multi-dimensional source model can be reduced to a one-dimensional model which reserves most of the statistical properties of the original source model. This greatly reduce the complexity of the link-level problem while yield approximate results reasonably close to the simulation results. The reduced-load method for solving the network problem will be feasible only when the link-level problem can be solved promptly.

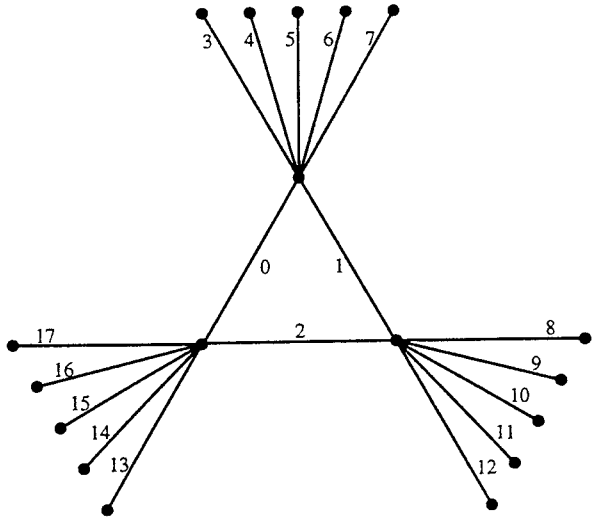

Figure 3: A multi-link ATM loop network.

In order to validate the approximation, the analytical results are compared with simulation results for several examples of ATM network and switch architectures with and without feedback effect. An example of multi-link ATM loop network with feedback is shown in Figure 3. There are three types of on-off sources in the traffic, representing data (type 1), voice (type 2), and video (type 3), respectively. Each type of the on-off source has different peak access rate γ, mean time at the ON state T_{ON}, and mean time at the OFF state T_{OFF}. In Table 2, simulation results are compared with results obtained via the approximation algorithm. Values of v_i's represent the number of each type of on-off source along each path, respectively. $\hat{R}_{Li}$'s denote the packet loss probability of of each type of on-off source along each path, respectively. $\overline{D}$'s denote the packet delay time along each path, respectively. In order to reduce the size of the table, we assume that the traffic through the loop network is symmetric geometrically. The algorithm also work well for general traffic patterns. From the numerical results, the accuracy of the approximation is validated and the efficiency of the approximation is also established. For the network examples we used, while simulation requires several hours of CPU time to give the end-to-end QoS, the approximation takes less than several minutes to obtain close approximate results.

$T_{ON1} = 40\ ms$, $T_{OFF1} = 60\ ms$, $r_1 = 16\ Kbps$,
$T_{ON2} = 400\ ms$, $T_{OFF2} = 600\ ms$, $r_2 = 64\ Kbps$,
$T_{ON3} = 386\ ms$, $T_{OFF3} = 765\ ms$, $r_3 = 1.14\ Mbps$,
Packet Size = 512 bits, Buffer Size = 50 packets,
Link Capacity = 30 $Mbps$ for Link 0-2,
Link Capacity = 3 $Mbps$ for Link 3-17.
A: analysis S: simulation

Path	v_1	v_2	v_3		$\hat{R}_{L1}$	$\hat{R}_{L2}$	$\hat{R}_{L3}$	$\overline{D}(ms)$
3-1-2	6	7	5	A	.0371	.0386	.0766	2.85
				S	.0543	.0468	.0892	2.94
4-1-2	4	5	7	A	.1024	.1050	.1711	4.73
				S	.0913	.0868	.1475	4.32
5-1-2	5	7	4	A	.0184	.0193	.0434	1.77
				S	.0270	.0170	.0449	1.84
6-1-2	3	9	6	A	.0730	.0753	.1324	3.85
				S	.0698	.0850	.1502	4.04
7-1-2	8	6	7	A	.1067	.1094	.1771	4.74
				S	.1266	.1015	.1722	4.83

Table 2: Performance for the ATM loop network.

PERFORMANCE ANALYSIS FOR VIDEO MULTICASTING NETWORKS

In multicasting, each source communicates with a group of receivers who are distributed over the network. Receivers may be able to share part of their connections. In addition, receivers may have different service requirements and/or capabilities. As a result, subband coding is usually used for encoding the source. Receivers who have a large access bandwidth will have a better quality of service (e.g. smaller dropping probability). However, a large access bandwidth will also introduce a higher call blocking probability because the system will become heavily loaded. The exact value for the end-to-end blocking probability is extremely difficult to obtain even for small networks [7]; only approximations are feasible. To this end we have developed an efficient approximation technique[8] for evaluating the performance of video circuit-switched multicasting networks in terms of end-to-end call blocking probabilities and packet dropping probabilities.

The technique is based on the reduced load method and is different than the one developed for the packet- or cell-switched networks above. The probabilities of blocking at different links of the same path are first assumed to be independent, and the link-level blocking

probabilities are approximated. Next, the traffic load of every link of a path is reduced by the blocking of all up-stream and down-stream links of that path. The reduced traffic loads are then used to update the link-level blocking probabilities. This leads to a nonlinear fixed-point problem, which can be solved by repeated substitutions. Approximations to the end-to-end blocking probabilities then follow immediately. Numerical results are obtained for the network shown in Figure 4. The traffic loads are varied and the average blocking and dropping probabilities are plotted in Figure 5. The error bars correspond to the 95% confidence intervals obtained by the Monte-Carlo method. From the graphs, we found that the approximations obtained are accurate.

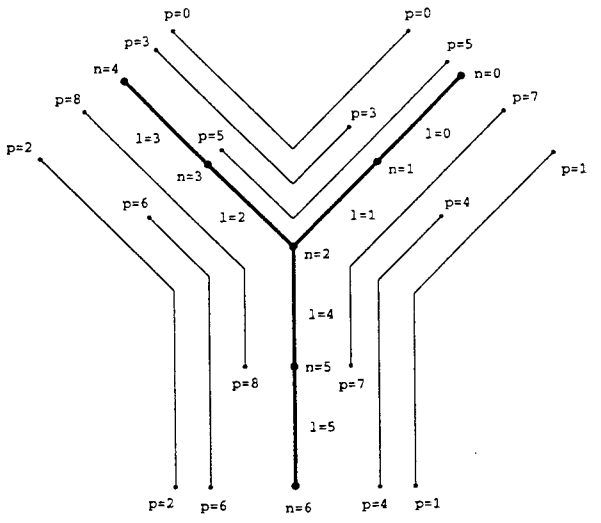

Figure 4: A sample multicasting network.

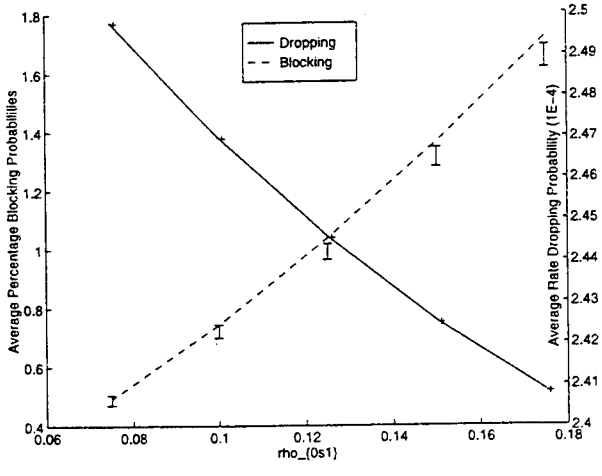

Figure 5: Tradeoff between call blocking and rate dropping probabilities.

RESOURCE ALLOCATION FOR ATM NETWORKS

Traces of actual traffic measurement data are used to perform event-driven simulation. Three kinds of shuffling and smoothing methods are also used to change the characteristics of the actual data traces in different ways in order to test the relative importance of their various statistical features on ATM network performance. From the simulation results, allowable resource allocation regions set by the three fundamental constraints of ATM networks and dominant regions for the three statistical features of the traffic source can be identified [9]. These provide useful insights into the performance prediction and resource allocation of ATM networks. The performances of fixed and three heuristic dynamic link bandwidth allocation schemes [10] on a FCFS queue with finite buffer size, bursty traffic input, and service rate equal to the allocated link bandwidth are also compared. The queueing model considered is shown in Figure 6 and the allowable region with fixed link bandwidth allocation scheme is shown in Figure 7.

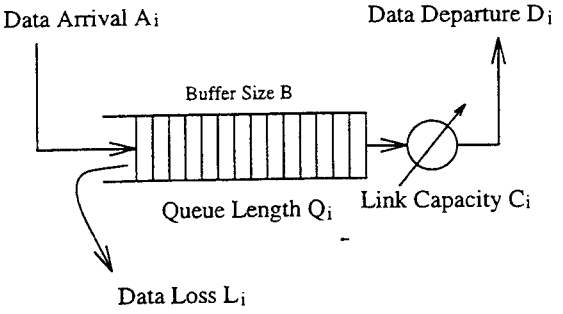

Figure 6: An ideal queueing model.

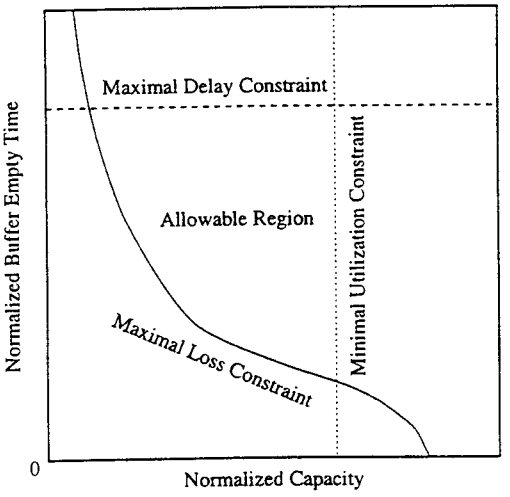

Figure 7: Allowable resource allocation region.

Figure 8 shows the performance under a dynamic scheme based on measurement of the data arrival rate A_i to maintain a target utilization factor. The data trace used is the Ethernet external packet traffic (Trace 4) which is very bursty as can be judged from its high peak/mean ratio and Hurst parameter. We assume that the allocated bandwidth will be updated periodically at the end of each allocation update interval, which is an integer multiple U of the measurement interval T. The buffer size B is the value normalized to the mean data arrival rate of the data trace. From the figure we can see that dynamic scheme perform better than fixed scheme in that the same ρ can be achieved with smaller p or higher ρ can be achieved with the same p, given that B is fixed. Also obvious is that higher allocation update rate (i.e., smaller U) will results in better performance of the dynamic schemes. By using dynamic scheme with a high allocation update rate, two orders of magnitude improvement in p over the fixed scheme can be achieved given the same values of ρ and B. This is a remarkable advantage of the dynamic allocation schemes. Analytical solving methods using MMPP traffic models to obtain the network performance under dynamic link bandwidth allocation schemes are also developed.

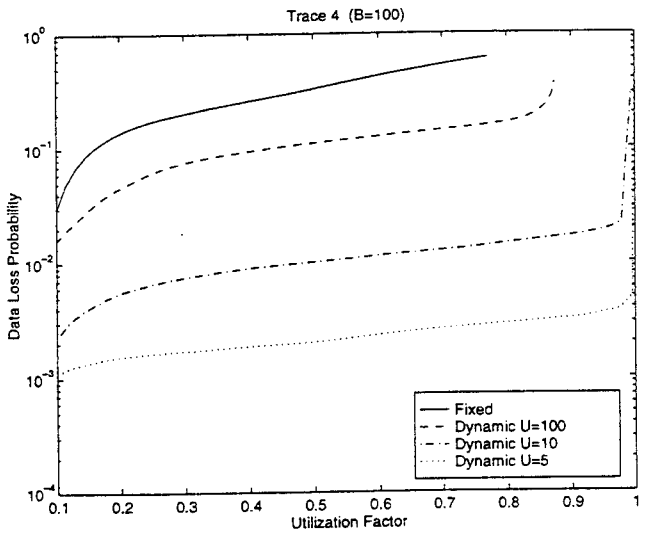

Figure 7: Performance of the dynamic link bandwidth allocation scheme.

CONCLUDING REMARKS

Possible extension to this work include: (1) Use the approximate end-to-end performance measures instead of exact expressions or simulations as objective functions in control/protocol optimization and sensitivity analysis. (2) Develop optimized dynamic schemes for admission control and bandwidth allocation of multimedia traffic in ATM networks. (3) Buffer management of ATM switches with multi-media traffic. (4) Routing of virtual paths in ATM networks.

The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Laboratory or the U.S. Government.

References

[1] J. Beran, R. Sherman, M. S. Taqqu, and W. Willinger, "Long-Range Dependence in Variable-Bit-Rate Video Traffic," *IEEE Trans. Commun.*, vol. 43, no. 2/3/4, pp. 1566-1579, Feb./March/April 1995.

[2] W. E. Leland, M. S. Taqqu, W. Willinger, and D. V. Wilson, "On the Self-Similar Nature of Ethernet Traffic (Extended Version)," *IEEE/ACM Trans. Networking*, vol. 2, no. 1, pp. 1-15, Feb. 1994.

[3] V. Paxson and S. Floyd, "Wide-Area Traffic: The Failure of Poisson Modeling," *Proc. ACM SIGCOMM '94*, pp. 257-268, Sep. 1994.

[4] C.-H. Chou and E. Geraniotis, "Effects of Traffic Higher-Order Statistics and Time Scales on Resource Allocation of ATM Networks," to be submitted.

[5] C.-H. Chou and E. Geraniotis, "Modeling of Long-Range Dependent Traffic in ATM Networks Using Heterogeneous On-Off Sources", to appear in *Proc. ICT*, Melbourne, April 97.

[6] C.-H. Chou and E. Geraniotis, "Efficient Computation of End-to-End Performance Measures for Multi-Link ATM Networks with Multi-Media Traffic," *Proc. IEEE INFOCOM ,95*, pp. 170-178, April 1995.

[7] F. Kelly, "Blocking Probabilities in Large Circuit-Switched Networks," *Adv. in App. Prob.*,, vol. 18, pp. 473-505, 1986.

[8] W.-C. Chan, and E. Geraniotis, "Limiting the Access Bandwidth of a Video Source: Model and Performance Analysis," *Proc. ICCCN*, pp. 546-553, 1995.

[9] C.-H. Chou and E. Geraniotis, "Performance Prediction and Resource Allocation for Long-Range Dependent Traffic in ATM Networks," *Proc. CISS*, pp. 198-204, March 1996.

[10] C.-H. Chou and E. Geraniotis, "Dynamic Link Bandwidth Allocation for Long-Range Dependent Traffic in ATM Networks", *Proc. IEEE ATM '96 Workshop*, San Francisco, August 1996.