Economides, A.A. & Kona, A.: **Adaptive routing of video traffic in ATM networks.** *Proceedings 6th Joint European Networking Conference*, pp. 131.1-131.7, TERENA 1995.

Adaptive Routing of Video Traffic in ATM Networks

Anastasios A. Economides and Altin Kona

<economid@macedonia.uom.gr>

Abstract

Asynchronous Transfer Mode (ATM) has been accepted as the transport mode for implementing Broadband Integrated Services Digital Networks (B-ISDN). In these networks different traffic classes with widely varying characteristics and conflicting service requirements will be statistically multiplexed and share common switching and transmission resources. It is important to control the traffic and the network resources to provide guaranteed levels of Quality of Service (QOS) for these different traffic classes.

In this paper we consider the adaptive routing as a traffic control mechanism in ATM networks with video and data users. Video-calls may have non-preemptive priority or a weighted round-robin scheduling over data-calls. Every video-call is independently routed over the path that had the minimum average video-cell delay. Similarly, every data-call is sent through the path that had the minimum average data-cell delay. Simulation results illustrate that adaptive routing responds to traffic variations and achieves extremely improved performance compared to static routing. Furthermore, the weighted round-robin service disciplines result in much better performance for the data-calls than the priority service discipline. Performance measures that are compared are the average cell delay, blocking probability of cells, and percentage of cells with delay above a threshold.

Keywords:

Adaptive routing, ATM networks, bandwidth allocation, service discipline, video traffic.

I. Introduction

ATM is the transmission, multiplexing and switching technique chosen for the B-ISDN. It can support and integrate many different types of services with different traffic characteristics and QOS requirements. This makes ATM particular appropriate for use with multimedia services. ATM networks are connection-oriented networks and require a virtual call to be established for each source-destination connection. The routing decisions are made at the virtual call establishment phase and all cells belonging to this virtual call follow the same path through the network. This allows each communication path to be customised based on the desired QOS requirements, for example: cell delay, cell delay variation, cell loss ratio e.t.c.

Adaptive routing in ATM networks decides on which route to send a new virtual call according to real-time traffic measurements collected at the network nodes [3,4,5,6,14,15].It allocates a new virtual call over the route that has the minimum length, where the length of the route may be the sum of the link lengths on this route. Since the decisions are on-line, adaptive routing can cope with the real-time traffic uncertainty and fluctuations.

Adaptive routing is also used in traditional packet-switched and telephone networks, such as ARPANET, TYMNET, TRANSPAC, TELENET, TSMR (AT&T), DCR (BNR), STAR (CNET), DAR (BT) and STR (NTT) [8,11]. The link length in ARPANET is the average packet delay over this link, computed over a 10 sec interval [8,9]. Every 10 seconds, each node computes the average delay on each outgoing link and sends it to all other nodes [1,17]. The link length in TYMNET is a function of the link capacity, the type of transmission facilities, the link load condition and the type of call that made the route generation request [9]. The link length in GTE TELENET is the ratio of the number of active calls currently supported on the link and the link capacity [7,9]. The link length in TRANSPAC is based on the link utilisation and on the number of calls routed on the link averaged over a 10 sec period [2,8]. The link length in the Internet is the packet delay [12,13].

Despite the tremendous research activity in the control of ATM networks, adaptive routing has not received much attention as a traffic control mechanism. The purpose of this paper is to investigate the impact of the adaptive routing combined with a priority or weighted round-robin scheduling on the QOS requirements of video and data users. Note that Priority Assignment Control [16], and Asynchronous Time Sharing [10] have been also proposed as QOS control methods. In section II, we describe our simulation model. We consider an ATM network with video and data traffic. Each user in a traffic class is modeled with a 3-level model (call, burst, cell). Each call from a traffic class is routed and served differently than calls from other traffic classes. In section III, we report the simulation results for adaptive and static routing with weighted round-robin and priority service disciplines. We measure the average cell delay, the blocking probability of cells and the percentage of cells with excess delay. In section IV, we conclude that adaptive routing provides extremely improved performance compared to static routing. Also, the priority service discipline produces large deterioration to the performance of the low priority traffic.

II. Simulation Model

In this section, we describe the simulated network, the traffic characteristics, the routing schemes, the service disciplines and the performance measures.

We consider an ATM network, where the information is packaged in cells of 53 bytes length (424 bits). Users from two different classes alpha and beta use two paths, each of 150 Mbps, to transmit their traffic to a destination (Figure). In front of each path, there are one buffer for class alpha cells and another buffer for class beta cells. Each buffer may store a maximum of 100 cells. Cells that find their corresponding buffer full are rejected.

While the transmission of class alpha and beta$ traffic is under way, a very high priority traffic of 100 Mbps starts transmission over the second path. The result is that users from class alpha and beta$ have now available for transmission a path of 100 Mbps and another path of 50 Mbps. We are interested in getting two answers:

1) how much improvement the adaptive routing offers over the static routing, and

2) what is the impact of the routing decisions and the service scheduling on the QOS requirements of the video and data traffic.

In class **alpha**, either 10, 20, 30, 40, 50 or 60 users simultaneously transmit video-calls. In class beta, 100 users simultaneously transmit data-calls.Each user in class alpha transmits video-calls of an exponential call duration with mean 1.875 sec and exponential interarrival time of calls with mean 5.625 sec (Table 1, Figure). Each video-call is divided into frames of fixed length F=62.5 msec.During each frame the user transmits at a peak rate of 10 Mbps, for a time interval $T^{video}{}_{ON}$ uniformly distributed in [10 msec, 40 msec] [10].

Each user in class beta transmits data-calls of an exponential call duration with mean 0.6 sec and exponential interarrival time of calls with mean 1.2 sec (Table 1, Figure). Each data-call alternates between an activity period, exponentially distributed with mean $T^{data}{}_{ON}$ 73 msec, and a silence period, exponentially distributed with mean $T^{data}{}_{OFF}$ 2 msec. During the activity period the user transmits at a peak rate of 64 Kbps.

In ATM networks the routing decisions are done at the call level. Each call is routed independently and all cells belonging to this call follow the same route [3,4,5,6]. So, the routing decisions affect a large amount of traffic. Therefore the impact of the routing decisions on the video and data traffic performance is severe. ATM networks should be designed to efficiently accomodate multiple traffic types with widely varying QOS requirements. For example, the end-to-end cell blocking probability is required to be less than 10^{-12} for HDTV, and 10^{-9} for remote procedure call. Also, the end-to-end cell delay is required to be less than 40 msec for HDTV, and 10 msec for remote procedure call. In this paper, we consider a simple ATM network and investigate the impact of the routing and of the service discipline on the average cell delay, the blocking probability of cells, and the percentage of cells with excess delay.

traffic level	video-user	data-user
mean call duration	1.875 sec	0.6 sec
mean interarrival time of calls	5.625 sec	1.2 sec
mean activity duration T_{ON}	[10 msec, 40 msec]	73 msec
mean silence duration T_{OFF}	62.5 msec T_{ON}	2 msec
peak bit rate during activity	10 Mbps	64 Kbps

Table 1. Traffic characteristics for each video-user and data-user.

We consider two routing schemes: static and adaptive routing. In the static routing scheme, the router has no information about the network state. Therefore the routing decisions are based on a priori knowledge about the network topology and capacities. The router only knows that there are two paths of 150 Mbps. It does not know that the second path is temporarily used by a very high priority traffic. So, it splits the traffic, half to the first path and the rest half to the second path.

In the adaptive routing scheme, there are two routers: one for the video-calls and another one for the data-calls. The video-router is informed about the average delay experienced by video-cells on each path during every 100 msec updating period. This information is sent to the video-router at the end of this updating period and becomes available to him after a feedback delay of 10 msec. When a new video-call arrives to the video-router, he sends it through the path that had the minimum average video-cell delay. Similarly, the data-router collects information about the average data-cell delay on each path every 100 msec time period. Then, he sends each newly arriving data-call through the path that had the minimum average data-cell delay. We distinguish the routing decisions among the video-calls and data-calls, because a path that seems good for a video-call may be very bad for a data-call.

We also compare five different service disciplinesfor serving cells from video-calls and data-calls: "1-1": cells from each class are served alternatively, one video-cell, then one data-cell, and so on.

"2-1": two video-cells are served, then one data-cell, and so on.

"5-1": five video-cells are served, then one data-cell, and so on.

"10-1": ten video-cells are served, then one data-cell, and so on.

"priority": every video-cell has non-preemptive priority over every data-cell.

For all service disciplines, when it is time to serve cells from a class but its corresponding buffer is empty, then cells from the other class are served. Summarizing, we have users from two different traffic classes with different characteristics, different routing decisions, and different service scheduling. Furthermore, we employ a 3-level traffic model for each user, adaptive routing, as well as weighted round-robin and priority service disciplines. Obviously, it is impossible to mathematically model and solve this problem. Thus, we are driven to simulation. In our simulation program, we consider all possible combinations of the following variables:

a) number of video-users (six cases): 10, 20, 30, 40, 50, 60,

b) service discipline (five cases): "1-1", "2-1", "5-1", "10-1", "priority",

c) routing scheme (two cases): adaptive, static.

In the next section, we present the simulation results.

III. Simulation Results

In this section, we analyse and discuss the simulation results. First, we compare the adaptive and static routing, then, the weighted round-robin and priority service disciplines.

III.A. Adaptive and static routing

In brief, the average cell delay and cell blocking in adaptive routing are much smaller than those in static routing. As the traffic load increases, this difference becomes huge. In heavy traffic, the static routing performs worst. Now, let us discuss the results more extensively.

First, for video-cells: Table 2 shows that the average video-cell delay in adaptive routing is always much smaller than that in static routing, especially in heavy traffic. For example, when there are 60 video-users, the average video-cell delay in adaptive routing is 25 times smaller than that in static routing. Also, as the number of video-users increases, the average video-cell delay increases slowly in adaptive routing, but very fast in static routing. For example, in adaptive routing, this delay for 60 video-users is only 2 times greater than that for 10 video-users.

On the other hand, in static routing, it is 25 times greater than that for 10 video-users

video-users	daptive 1-1"	tatic 1-1"	daptive priority"	tatic priority"
10	.99	7.39	2.99	.01
20	3.20	14.20	3.18	1.29
30	3.43	41.82	3.37	30.91
40	3.77	75.19	3.78	69.19
50	4.78	132.92	4.47	118.90
60	7.29	185.41	6.60	166.17

Table 2. The average video-cell delay μsec for adaptive and static routing with "1-1" and "priority" service disciplines.

ideo-sers	daptive 1-1"	tatic 1-1"	daptive priority"	tatic priority"
0		6873		3094
0		1068		3896
0		61013		94718
0	58	298543	02	233267
0	021	200957	919	881469

0	9701	637707	7949	897773

Table 3. The video-cell blocking probability (*10^{-8}) for adaptive and static routing with "1-1" and "priority" service disciplines.

Table 3 shows that the blocking probability of video-cells in adaptive routing is extremely smaller than that in static routing. There is no video-cell blocking in adaptive routing until there are 40 video-users. In static routing, video-cell blocking starts developing even for 10 video-users, and is becoming 2300 times larger than that in adaptive routing for 40 video-users. Note also, that the video-cell blocking is developing very fast, as the number of video-users increases.

video-users	daptive 1-1"	tatic 1-1"	daptive priority"	tatic priority"
.10		0.2614	0	.1218
20	0	1.9784	0	.8432
30	0.0038	7.2011	0.0031	3.8786
40	0.0517	2.5789	0.0485	9.5769
50	0.3339	0.5198	0.2362	16.4043
60	1.1420	6.7017	0.9061	22.5451

Table 4. The percentage of video-cells with delay over 50 μsec for adaptive and static routing with "1-1" and "priority" service disciplines.

For example, in adaptive routing with "1-1" service, the video-cell blocking for 60 video-users is 70 times greater than that for 40 video-users. On the other hand, in static routing with "1-1" service, for 60 video-users it is 334 times greater than that for 10 video-users.

Table 4 shows that the percentage of video-cells with delay over 50 μsec in adaptive routing is extremely smaller than that in static routing. In adaptive routing, no video-cell experiences delay over 50 μsec until there are 30 video-users. In static routing, some video-cells experience delay over 50 μsec even for 10 video-users. For 30 video-users, the percentage of video-cells experiencing delay over 50 μsec in static routing is 1900 times larger than that in adaptive routing. Similar to the video-cell blocking, the percentage of video-cells with excess delay starts growing very fast, as the number of video-users increases. For example in adaptive routing, this percentage for 60 video-users is 300 times greater than that for 30 video-users. In static routing with "priority" service, this percentage for 60 video-users is 185 times greater than that for 10 video-users.

For data-cells:

Table 5 shows that the average data-cell delay in adaptive routing is always much smaller than that in static routing, especially when video-cells have priority over data-cells. For example, when there are 40 video-users with "priority" service, the average data-cell delay in adaptive routing is 160 time smaller than that in static routing. Furthermore, as the number of video-users increases, the average video-cell delay increases slowly in adaptive routing, but very fast in static routing. For example, in adaptive routing with "priority" service, this delay for 60 video-users is only 16 times greater than that for 10 video-users. On the other hand, in static routing with "priority" service, this delay for 60 video-users is 384 times greater than that for 10 video-users.

video-users	daptive 1-1"	tatic 1-1"	daptive priority"	tatic priority"
10	.99	6.08	3.02	.75
20	3.14	6.41	3.28	2.39
30	3.26	6.78	3.67	286.46
40	3.40	7.09	6.08	944.28
50	3.52	7.45	15.79	2176.99
60	3.64	7.68	48.06	3741.37

Table 5. The average data-cell delay (μsec) for adaptive and static routing with "1-1" and "priority" service disciplines.

video-users	daptive" -1"	tatic 1-1"	daptive priority"	tatic priority"
10		0	0	890
20	0	0	0	6478
30	0	0	0	207930
40	0	0	1765	107318 3
50	0	0	9352	337616 4
60	0	0	38862	704098 4

Table 6. The data-cell blocking probability (*10^{-8}) for adaptive and static routing with "1-1" and "priority" service disciplines.

Table 6 shows that there is data-cell blocking only when the video-cells have priority over the data-cells. In that case, there is no data-cell blocking in adaptive routing until there are 40 video-users. However, in static routing, it starts developing even for 10 video-users, and it becomes 600 times larger than in adaptive routing for 40 video-users. Furthermore, we notice that as the number of video-users increases, it is developing very fast. For example, in adaptive routing with "priority" service, for 60 video-users, it is 22 times greater than that for 40 video-users. Also, in static routing with "priority" service, for 60 video-users, it is 2400 times greater than that for 40 video-users.

video-users	daptive" -1"	tatic 1-1"	daptive priority"	tatic priority"
10		0.00001	0	.0433
20	0	0.0001	0	.6724
30	0	0.0003	0.0034	3.0686
40	0	0.0006	0.0519	7.6422
50	0	0.0008	0.2918	13.2598
60	0	0.0008	0.9443	17.5552

Table 7. The percentage of data-cells with delay over 50 μsec for adaptive and static routing with "1-1" and "priority" service disciplines.

Table 7 shows that in adaptive routing, no data-cell experiences delay over 50 μsec until there are 30 video-users and "priority" service. On the other hand, in static routing, data-cells start experiencing delay over 50 μsec even for 10 video-users. For 30 video-users, the percentage of data-cells experiencing delay over 50 μsec in static routing is 900 times larger than that in adaptive routing. Similar to the data-cell blocking, the percentage of data-cells with excess delay starts growing very fast, as the number of video-users increases. For example, in adaptive routing with "priority" service, this percentage for 60 video-users is 278 times greater than that for 30 video-users. Also, in static routing with "priority" service, this percentage for 60 video-users is 405 times greater than that for 10 video-users.

Thus, we remark that adaptive routing provides extremely improved performance for both the video-calls and the data-calls.

III.B. Service discipline

The data-traffic is light compared to the video-traffic and it does not seriously affect the performance of the video-traffic. We find that the different round-robin service disciplines provide similar performance for the video-traffic. However, the average video-cell delay, video-cell blocking probability and the number of video-cells with excess delay in "priority" service are obviously smaller than those in the other disciplines (as expected), but this improved performance is not very large (Tables 2, 3, and 4). On the other hand, the effect of the "priority" service on the data-traffic is dramatic. Although the different round-robin service disciplines provide similar performance for the data-traffic, the "priority" service prevents the data-cells to be transfered efficiently.

Since, the different round-robin service disciplines provide similar performance, we only show the performance of the "1-1" and the "priority" service.

Table 5 shows that in adaptive routing with 60 video-users, the average data-cell delay with "priority" service is 13 times greater than that with "1-1" service. This difference worsen in static routing, where it is 487 times greater than that with "1-1" service.

Table 6 shows that there is not data-cell blocking with the "1-1" service, since the data-traffic is low enough to be served efficiently. However, the data-cell blocking probability with "priority" service is extremely high. This happens, because the video-cells have priority over the data-cells and monopolise the servers. So, the data-cells find their buffer full and are rejected.

Similarly, Table 7 shows that no data-cell experiences delay over 50 μsec in adaptive routing with "1-1" service. However, in adaptive routing with "priority" service, the percentage of data-cells experiencing delay over 50 μsec is large. In static routing, it is even worst, since this percentage with "priority" service may become 22000 times greater than that with "1-1" service.

Thus, under "priority" service, the data-users have severe performance degradation.

IV. Conclusions

In this paper, we investigate the adaptive routing as a traffic control mechanism in ATM networks with video and data traffic. We consider up to 60 video-users and 100 data-users that route their calls over a two path ATM network. We model each user with a 3-level traffic model. So, each video-user transmits video-calls (call level) with exponentially distributed interarrivall time of video-calls and video-call durations. During a fixed frame duration, each video-call transmits for an activity period uniformly distributed (burst level). The peak bit rate (cell level) during the activity period is constant. Similarly, a data-user transmits data-calls with exponentially distributed interarrivall time of data-calls and data-call durations (call level). Each data-call alternates between exponentially distributed periods of activity and silence (burst level). During the activity period, the peak bit rate is constant (cell level).

We employ adaptive routing by making routing decisions according to real-time measurements of the delay on the paths. The routers are periodically informed about the current status of the network paths and decide accordingly.

We introduce class-depended routing by distinguishing the routing decisions for each traffic class. Calls from each traffic class are routed independently from calls from the other class. So, video-calls are routed through the path that provides the minimum average delay for video-cells, while data-calls through the path that provides the minimum average delay for data-cells.

Each traffic class has not only different traffic characteristics, and is routed independently from the other class, but it is also served in a different way than the other class. We employ class-depended service discipline for scheduling the transmission of cells over each path. So, video-cells may have non-preemptive priority over data-cells, or served according to a weighted round-robin service discipline.

Subsequently, we develop a simulation program to model all these ideas and investigate the impact of the adaptive routing, the service discipline and the traffic load on the QOS requirements of the video and data calls.

In the simulation environment, the data-traffic is small compared to the video-traffic. So, for all non-priority service disciplines, the data-traffic is served efficiently. However, for "priority" service, the video-traffic monopolises the network and the data-traffic has serious performance degradation. Furthermore, this performance deterioration of the data-traffic is more intensive in static routing.

When the high priority video-traffic is heavy compared to the low priority data-traffic, then the high priority traffic monopolises the network resources and the low priority traffic cannot be served efficiently. When it is important to efficiently serve the data-traffic, the network administrator should use a more "controllable" service discipline, instead of strict priorities.

Employing a weighted round-robin service seems more appropriate. Then, depending on the traffic load and the QOS requirements, the administrator determines the weights in the service discipline to meet the QOS of the different traffic classes.

On the other hand, if it is not important to meet some strict QOS requirements for the data-traffic, but it is very important to achieve the best performance for the video-traffic, the administrator may use strict priorities.

Finally, adaptive routing always provides much smaller average cell delay, cell blocking probability, and percentage of cells with excess delay than the static routing. This improved performance is even better in heavy traffic, where adaptive routing performs extremely better than static routing.

V. References

[1]Amer, F. and lien, Y.-N. A survey of hierarchical routing algorithms and a new hierarchical hybrid adaptive routing algorithm for large scale computer communication networks. Proc. ICC'88, pp. 999-1003, 1988.

[2]Brown, P., Roumilhac, J. and Bonnard, P. A study of the TRANSPAC routing algorithm Teletraffic Science for New Cost-Effective Systems, Networks and Services, ITC-12, M. Bonatti (editor), pp. 1033-1039, Elsevier Science Publ. 1989.

[3] Economides, A.A., Ioannou, P.A. and Silvester, J.A. Decentralized adaptive routing for virtual circuit networks using stochastic learning automata Proc. of IEEE Infocom 88 Conference pp. 613-622, IEEE 1988.

[4] Economides, A.A., Ioannou, P.A. and Silvester, J.A. Adaptive Routing and Congestion Control for Window Flow Controlled Virtual Circuit Networks Proc. 27th Annual Allerton Conference on Communications, Control, and Computing pp. 849-857, Sept. 1989.

[5] Economides, A.A., Ioannou, P.A. and Silvester, J.A. Dynamic routing and admission control for virtual circuit networks Journal of Network and Systems Management Vol 2, No 2, 1995.

[6] Economides, A.A., Ioannou, P.A. and Silvester, J.A. Adaptive virtual circuit routing Computer Networks and ISDN Systems Vol. 27, 1995.

[7] Epelman, M. and Gersht, A. Analytical modeling of GTE TELENET dynamic routing Teletraffic Issues in an Advanced Information

Society, ITC-11, M. Akiyama (editor), Elsevier Science Publ. 1985.

[8] Gerla, M. Controlling routes, traffic rates, and buffer allocation in packet networks IEEE Communications Magazine Vol. 22, No. 11, pp. 11-23, Nov. 1984.

[9] Hsieh, W.-N. and Gitman, I. Routing strategies in computer networksIEEE Computer, pp. 46-56, June 1984.

[10] Hyman, J.M., Lazar, A.A. and Pacifici, G. Real-time scheduling with quality of service constraints, IEEE Journal on Selected Areas in Communications, Vol. 9, No. 7, pp. 1052-1063, Sept. 1991.

[11] Mase, K. and Yamamoto, Y. Advanced traffic control methods for network management, IEEE Communications Magazine, pp. 82-88, Oct. 1990.

[12] Moy, J. OSPF: next generation routing comes to TCP/IP networks, LAN Technology, pp. 71-79, April 1990.

[13] Narten, Th. Internet routing, Proc. ACM Communication Architectures and Protocols, pp. 271-282, 1989.

[14] Onvural, R.O. Asynchronous Transfer Mode Networks: Performance Issues, Artech House, 1994.

[15] Onvural, R.O. and Nikolaidis, I. Routing in ATM networks, Proceedings of TriComm '92: High-Speed Communication Networks, H. Perros (editor), pp. 139-150, Plenum Press 1992.

[16] Takagi, Y., Hino, S. and Takahashi, T. Priority assignment control of ATM line buffers with multiple QOS classes, IEEE Journal on Selected Areas in Communications, Vol. 9, No. 7, pp. 1078-1091, Sept. 1991.

[17] Zinky, J., Vichniac, G. and Khanna, A. Performance of the revised routing metric in the ARPANET and MILNET, Proc. MILCOM, pp. 219-224, IEEE 1989.

Author Information

Anastasios A. Economides was born and grew up in Thessaloniki, Greece. He received the Diploma degree in Electrical Engineering from Aristotle University of Thessaloniki, in 1984. After receiving a Fulbright and a Greek State Fellowship, he continued for graduate studies at the United States. He received a M.Sc. and a Ph.D. degree in Computer Engineering from the University of Southern California, Los Angeles, in 1987 and 1990, respectively. During his graduate studies, he was a research assistant performing research on routing and congestion control. He is currently an Assistant Professor of Informatics at the University of Macedonia, Thessaloniki. His research interests are in the area of Performance Modeling, Optimization and Control of High-Speed Networks. He is a member of IEEE, ACM, INFORMS, EPY, TEE.

Altin Kona was born in Albania. He graduated in Informatics from the Faculty of Natural Sciences, University of Tirana, Albania, in 1992. He hold visiting positions in Germany, Italy and Greece. During 1992-93, he was with the Institute of Informatics and Applied Mathematics. Currently, he is a Lecturer at the University of Tirana.