21
Emulation of Traffic Congestion on ATM Gigabit Networks

J. Domingo-Pascual[*]*, A. Albanese, and W. Holfelder*[†]
International Computer Science Institute
1947 Center St., Suite 600, Berkeley, CA 94704-1198, USA
Phone: + 510 643-9153   Fax: + 510 643 7684
E-mail: jordid@ac.upc.es, {aa, whd}@icsi.berkeley.edu

Abstract

The deployment of gigabit networks and broadband services has started to support multimedia applications; however, these gigabit networks are rarely saturate since only a few applications are able to stress the network. We consider a future scenario where the use of multimedia applications, such as audio and video teleconferencing in a multi-user environment, is expected to grow rapidly. Therefore, both customers and network providers need to foresee the performance and behavior of the network and applications in this scenario. From the customer's point of view, it is important to develop procedures to perform traffic measurements and to be able to test the local ATM equipment. In this paper we propose a method for introducing heavy load into an ATM switch and at the User Network Interface (UNI) to study the performance and forecast evolved scenarios. In the experiments we used local equipment (ATM switch and workstations), local network management applications and diagnostics software. The emulated load is generated in a workstation, introduced into the ATM switch and intensified by replicating and re-circulating the cells. The method presented is an easy and affordable way to test performance and an alternative to traffic modeling. Several experiments have been performed and the measurements obtained are presented.

Keywords

ATM, traffic congestion, switch performance, traffic emulation.

[*] Funded by the Research Directorate of Catalonia (DGR) under contract BE94/3-590, and the Spanish Ministry of Education (CICYT) under grant number TIC92-1289-PB.

[†] Funded by a German Academic Exchange Service doctoral scholarship under grant number 516 402 5384 from HSP II/AUFE program resources.

R. Puigjaner (ed.), *High Performance Networking*

1 INTRODUCTION

ATM switching equipment in customer premises networks may have unacceptable performance under heavy loads; this may reduce the life-time of the equipment, in the sense that it has reached its limitations. Users need tools to foresee and deal with the problem of outgrowing the equipment deployed when broadband applications become widely used .

It is also important that these tools should be able to emulate high-load scenarios using low-cost techniques. In this paper we use traffic emulation with real existing applications for loading the switch, instead of random or model-driven traffic generation. When one cell is lost, the remaining cells that form the packet (MAC PDU, IP packet, etc.) are useless because the receiving entity will discard the whole packet. In a heavy-load environment cell losses may occur, causing long packets to have a smaller chance of being successfully transmitted than short packets.

In this paper we analyze the impact that cell losses have on various size packets under heavy network load. Section 2 presents the ATM trial network where the experiments were performed, including a description of the ATM cell relay service and local equipment used, and introduces the method for traffic emulation. Sections 3, 4, and 5 describe and show, for three experiments, the configuration set-up and the results obtained. Conclusions are presented in Section 6.

2 SET-UP OF THE EXPERIMENT

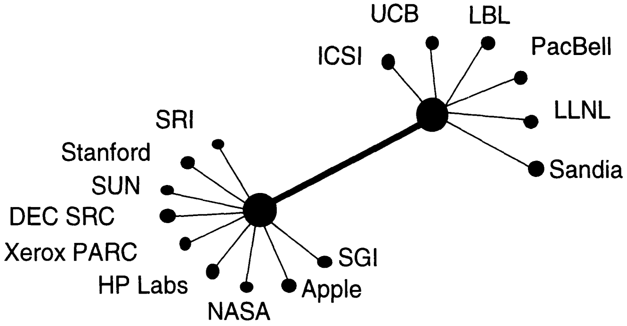
This section is divided into three sub-sections. The first one contains the description of the ATM cell relay service available in the ATM gigabit network. The second one is focused on the description of the local ATM equipment used and its configuration. The third sub-section briefly presents the traffic emulation method proposed.

2.1 ATM Cell Relay Service

The Bay Area Gigabit Network (BAGNet) has recently been installed to connect 15 computing research organizations located in the San Francisco Bay Area; the organizations have obtained support for ATM SONET services from CalREN (California Research and Education Network), a foundation established by Pacific Bell. The testbed participants, which include many of the government laboratories, research universities, research institutes, and technology companies in the Bay Area, are as follows: Apple Computer, Digital Equipment Corporation (Palo Alto Systems Research Center), Hewlett-Packard Laboratories, International Computer Science Institute, Lawrence Berkeley Laboratory, Lawrence Livermore National Laboratory, NASA Ames Research Center, Pacific Bell, Sandia National Laboratories, Silicon Graphics, Inc., SRI International, Stanford University, Sun Microsystems, Inc., University of California, Berkeley (Computer Science Dept. - Tenet Group), Xerox Palo Alto Research Center (Figure 1).

This network will be used for a number of collaborative multimedia applications, the first of which to be implemented is a “teleseminar” facility. This network and its applications will create an information highway within the Bay Area that will serve as a model for the national “information superhighway”.

Service Trial: BAGNet (CalREN)



- ATM service 155 Mbps (SDH)
- Applications: multi-party collaborations (initially teleseminar)

Figure 1 The Bay Area Gigabit Network.

The network offers an access rate of 155 Mbps. (SONET OC-3c) and the service provided is on a "best effort" basis; hence there are no QoS guarantees. All ATM drivers at the hosts use the AAL5 (non-assured mode) and LLC/SNAP encapsulation as described in ITU Recommendation I.363 and Heinanen (1993). The "Classical IP over ATM" approach defined in Laubach (1994) is used. Also, Heinanen (1993) and Atkinson (1994) will be followed. All hosts within the 15 sites form a Logical IP Sub-network (LIS), as defined in Laubach (1994). Each workstation has an IP address assigned within the BAGNet Sub-network (192.6.28.xxx), and it is associated with a fixed Permanent Virtual Circuit (PVC). Permanent VPs are used because signaling is not yet implemented by manufacturers and the testbed participants use equipment from different vendors. However, signaling will be available in the future implementing the ATM Forum UNI 3.1.

Because of PVC table limitations within the switch of the service provider each site can only connect up to four workstations. These may be connected either directly or through a local ATM switch. Each workstation is connected to every other workstation in a fully meshed topology through a point-to-point PVC; that makes 3540 PVCs (4 hosts/site * 59 connections/host * 15 sites). Furthermore, there is one point-to-multipoint PVC from each station to all the other hosts to provide an underlying multicast infrastructure. That makes 60 one-to-59 PVCs. This means that each workstation has 59 bi-directional point-to-point PVCs, 59 incoming uni-directional point-to-multipoint PVCs and 1 outgoing uni-directional point-to-multipoint PVC.

2.2 Local ATM Equipment

The infrastructure at ICSI consists of a local ATM switch switch (SynOptics Lattiscell 10114-SM) with 16 ports at 155 Mbps. (14 multi-mode fiber and 2 single-mode fiber ports), providing ATM all the way to the workstation. For these experiments we use two SUN SPARCstation 10s and one SUN SPARCstation 2, with their ATM card (SynOptics SBUS Adapter 975-192). The workstations are attached directly. Additionally, a fiber loop is used to connect ports 4 and 5. Port 15 is connected to the ATM network UNI provided by the carrier using a single-mode fiber. Figure 2 presents this scenario.

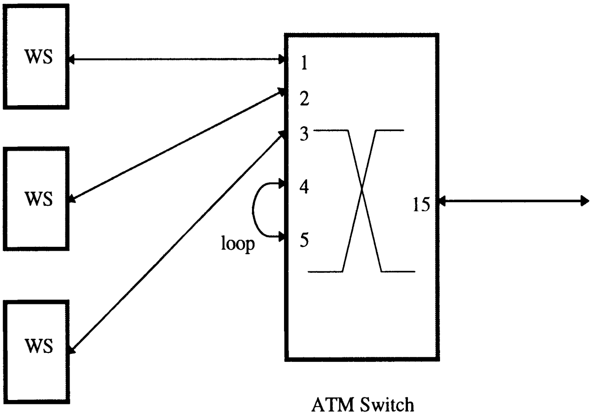


Figure 2 Scenario of the experiments.

The Network Management Application (SynOptics, 1994a) for the SynOptics switch is an extension to the SunNet Manager (Sun Microsystems, 1993a, 1993b, 1993c) and uses the Connection Management Services application (SynOptics,, 1994b). Network Management Application (NMA) allows the set-up of the connections for the ATM switch, both point-to-point and point-to-multipoint, to measure the bandwidth used and buffer occupancy for each port, and to retrieve the cell counters of transmitted, received, and dropped cells per each port and VC. NMA and CMS communicate with each other using SNMP on the Ethernet LAN. Due to software installation requirements CMS runs on one machine while NMA runs on a different one. In fact, NMA runs on one of the ATM workstations, even though it does not need to be connected to the ATM network directly and it would be better if it ran on a different workstation (Figure 3).

2.3 Traffic Emulation Method

The goal is to generate emulated traffic load in the network so as to make it as "realistic" as possible. Therefore, we use tools generating TCP and UDP traffic rather than random ATM traffic generators.

For our performance measurements, both at the cell level and the application level, we use the capabilities offered by the software tools. We are interested in results related to the traffic scenario we will have in future gigabit networks when many applications with a large number of users will run at the same time.

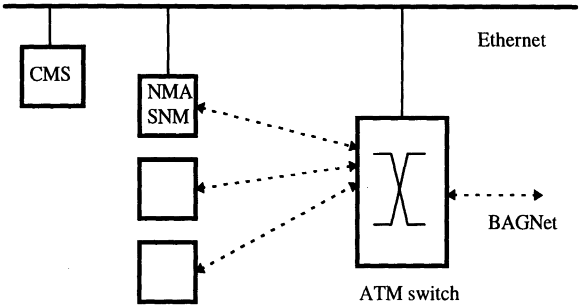


Figure 3 Configuration of the Network Management Applications.

The goal of the experiments is to study the effect of heavy traffic load on particular connections, taking into account both cell and packet losses. In particular we want to see the effects on the behavior of the application when cells are dropped due to local congestion at the input/output ports.

3 EXPERIMENT 1: UDP TRAFFIC BETWEEN LOCAL HOSTS

3.1 Set-up of the experiment

For the experiment we disconnected the local ATM network from the service in order to avoid external traffic. Workstations 1 and 2 were connected with a point-to-point PVC that goes through the fiber loop between ports 4 and 5. In Figure 4, the dotted line represents the normal route connecting the workstations attached to ports 1 and 2. For our experiment, however, we re-routed this connection as shown with the solid lines in Figure 4.

Workstation 3 generates the background traffic to load the switch. For this purpose, we established a uni-directional point-to-multipoint connection from port 3 to ports 4 and 5 (flows *a* and *b* respectively, in the figure). Each outgoing PVC at port 4 is looped back, through the external fiber loop, as an incoming PVC at port 5; from there another PVC sends the traffic back to port 4, closing the loop. The same applies to the other direction (flow *b*), from port 5 back to port 4. We can accomplish various load situations by changing the number of loops.

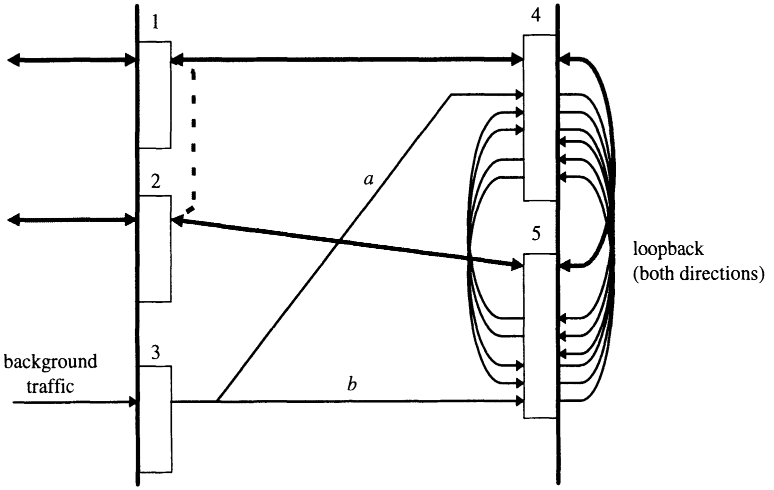


Figure 4 Connection set-up of the ATM switch for experiment 1.

With these two sets of loops (flows *a* and *b*) we replicate and re-circulate the traffic entering port 3, so that we load both the transmitting and receiving buffers at ports 4 and 5. Flow *a* feeds the loops that load the transmitting buffer at port 4 and the receiving buffer at port 5. Thus, it merges with the cells that go from workstation 1 to workstation 2. Flow *b* feeds the loops loading the transmitting buffer at port 5 and the receiving buffer at port 4, merging with the cells going from workstation 2 to workstation 1. The bi-directional connection between workstations 1 and 2, while crossing ports 4 and 5, will experience some cell losses depending on the congestion introduced into the re-circulating loops.

The background traffic from workstation 3 was generated using a testing tool called ttcp sending UDP packets. The traffic is sent to an address associated with the point-to-multipoint PVC at port 3. Depending on the load of the workstation, we were able to generate background traffic of up to 20 Mbps. (on a SPARC 2). Since these cells are replicated and then re-circulated in the loops, the number of loops determines the actual load. However, the throughput does not increase linearly with the load because some cells are dropped at the buffers of ports 4 and 5 due to buffer overflow. If we introduce 20 Mbps into the background traffic loops the actual throughput at the output port reaches about 120 to 134 Mbps. Adding more loops would not increase the throughput but only the number of cells dropped at the transmitting buffers. Depending on how much traffic we send from workstation 3 we can use more or fewer loops to generate a given cell loss ratio.

3.2 Results of experiment 1

We sent batches of 100 "pings" with different packet lengths from workstation 1 to workstation 2. Workstation 3 sent UDP traffic to the loops where we could measure 0.6% cell loss from port 4 to 5 and 1.2% cell loss from port 5 to 4. This impairment is due to the fact that we set up a different number of loops in each direction. We make the measurements with both re-circulating loops active (*a* and *b*), which means that we have losses in both directions, affecting both the ping requests and the ping responses. We repeated the measurements with only one direction re-circulating loops active, so that we had losses in only one direction.

In Table 1 we can observe that for packets of 512 bytes we get 70% packet loss if both traffic loops are active and 62% if only one is active. This suggests that most of the losses correspond to ping request packets and the other half are ping response packets. For packets of 1024 bytes almost all the losses are produced in one direction; with one direction traffic loop active 69% of the packets are lost, while with two direction traffic loops we obtain 73% lost packets. Long ping request packets are lost when congestion occurs at the output buffer.

Table 1 Ping packets lost between icsibag1 and icsibag2

Ping packet length (bytes)	*% of packets lost loops in both directions 0.6% / 1.2% cell losses*	*% of packets lost loops in one direction 1.2% cell losses*
64	35 %	23 %
128	62 %	50 %
256	50 %	51 %
512	70 %	62 %
1024	73 %	69 %
2048	86 %	75 %
4096	---	93 %

4 EXPERIMENT 2: UDP TRAFFIC BETWEEN LOCAL AND REMOTE HOSTS

The second experiment consisted of sending 'ping' packets to an external host, so that they traverse the ATM WAN network. The connection set-up for the switch is shown in Figure 5. As in experiment 1, we sent batches of 100 ping requests with different packet lengths.

4.1 Set-up of the experiment

We have a PVC for each point-to-point connection between hosts. In this experiment, the PVC that connects 'icsibag1' (the local workstation) with 'bag1' (the remote workstation) has VCI 92. This means that the VCI used at the UNI (port 15) is number 92. At the local host, the IP address of the remote host is bound to the corresponding PVC (port 1 and VCI 92, in our case). The connection set-up for the switch establishes the following path: port 1-VCI 92, port 4-VCI 92, external loopback, port 5-VCI 92 and port 15-VCI 92. This is a bi-directional point-to-point connection.

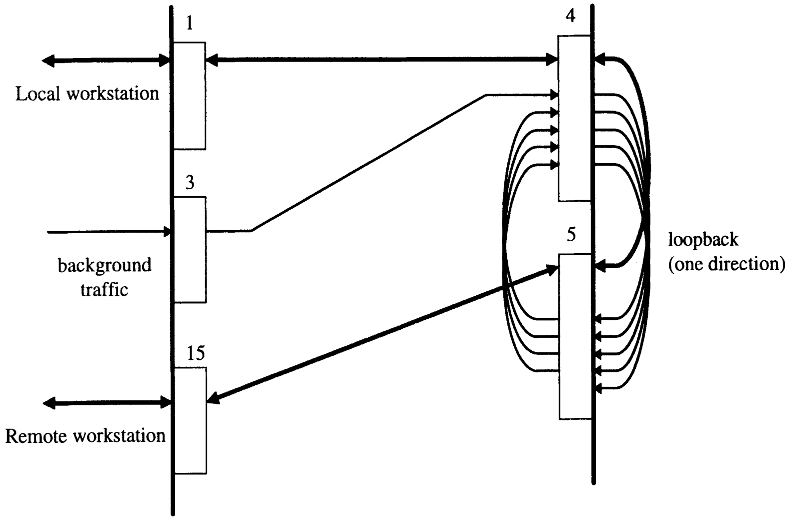


Figure 5 Connection set-up of the ATM switch for experiment 2.

4.2 Results of experiment 2

Using the NMA we monitor the counters for the cells received at ports 1 and 5 for VCI 92 (which corresponds to the cells of the ping requests), as well as cells received at ports 15 and 4 for VCI 92 involved in the reverse path (ping responses). This measurement allowed us to determine exactly where the cells were lost. Table 2 shows the measurements and the results obtained.

Cells received at port 1 (VCI 92) were originated at the local workstation, therefore they correspond to the 100 ping request packets. For a ping packet of 64 bytes three ATM cells are sent. The number of cells for each 64 byte ping packet is equal to the number of data bytes (64) plus the ping header (8), plus the IP header (20), plus the LLC/SNAP encapsulation (18), and the trailer byte of the AAL5 PDU. This is a total of 111 bytes which corresponds to 3 ATM cells (with 48 byte payload). For a packet length of 2048 bytes, each packet fills 45 cells because each ping message is segmented into two ICMP packets with their respective IP and LLC headers and the AAL5 trailer byte; so, the total number of cells sent during the experiment was 4500. These cells experience congestion when arriving at port 4, where they merge with the background traffic. Because of this, some cells are dropped at the transmitting buffer at port 4, and hence the number of cells arriving at port 5 (VCI 92) is smaller than the number of cells received at port 1. The difference is the number of cells lost at port 4.

For all packet lengths, cells entering through port 15 (VCI 92) are received by port 4 (VCI 92) and transmitted to workstation 1. We can observe that the number of cells received in each case is a multiple of the cells per packet. For instance, ping packets of 64 bytes are sent in 3 cells; 141 cells are received from the remote host, which means that 47 ping responses are received, and that 100 - 47 = 53 packets are lost. For 100 ping request packets of 1024 bytes, 33 ping response packets (23 cells each) are received. This agrees also with the statistics given by the ping program of 67% packets lost. In this way, we can keep track of the cells lost when traversing the loops.

Table 2 Cells sent and received when transmitting 100 pings from icsibag1 (local) to bag1 (remote)

Ping packet length	*64 bytes*	*256 bytes*	*512 bytes*	*1024 bytes*	*2048 bytes*
Cells at Rcv buffer port 1	300	700	1200	2300	4500
Cells at Rcv buffer port 5	254	565	981	1796	3069
Cells lost at port 4	46	135	219	504	1331
Cells at Rcv port 15	141	350	552	759	396
Cells at Rcv buffer port 4	141	350	552	759	396
Ping statistics: packets lost	53%	50%	54%	67%	91%

Also, we monitored the number of cells transmitted, the number of cells dropped at the transmitter buffer, and the number of cells received for ports 1, 3, 4, 5, and 15. Port 3 gives us the background traffic generated. The ratio between the cells received at port 3 (from workstation 3) and the total number of cells transmitted at port 4 gives us the multiplying factor of the loops, that is, how many times cells are multiplied when re-circulating the loops. On the other hand, the ratio between the number of cells dropped at the transmitting buffer at port 4 and the total number of cells transmitted at port 4 gives us the cell loss ratio (see Table 3).

Table 3 Overall cell loss ratio for the ping connection between icsibag1 and bag1 at port 4

Ping packet length:	*64 bytes*	*256 bytes*	*512 bytes*	*1024 bytes*	*2048 bytes*
Cells transmitted	33 905 219	33 844 868	34 384 145	34 088 408	34 133 629
Cells dropped	362 744	394 282	391 680	377 339	375 922
Cell loss ratio	1%	1.1%	1.1%	1.1%	1.1%
Ping statistics: packets lost	53%	50%	54%	67%	91%

For experiment 2 the background traffic generated by workstation 3 is between 23 and 25 Mbps. and it is multiplied by 5.6 to 5.7 times when going through the loops, giving an overall traffic between 132 and 144 Mbps. The cell loss ratio is about 1.1% as shown in Table 3. It must be taken into account that the cells transmitted at port 4 include the point-to-point connection and the background traffic loops, and that the dropping of cells affected both of them. The duration of each measurement was not exactly the same, and the traffic generated by workstation 3 is not absolutely constant, which is why the total number of cells through port 4 is not strictly the same for all cases.

Though the cell loss ratio is sustained in this experiment, the results show experimentally that cell losses during congestion intervals penalize long packets more seriously.

As a complement to experiment 2, we repeated the process with fewer cell losses. In order to do this, the traffic multiplying factor of the loops is 5, so that the overall load on port 4 does not exceed 125 Mbps. This gives a cell loss ratio of about 10^{-4} to $8*10^{-4}$, also measured as the ratio of cells dropped at port 4 and cells transmitted through it. The packet loss ratio experienced was: no losses for 64 byte packets, 13% losses for 2048 byte packets, 18% for 4096 byte packets and 39% for 8192 byte packets.

These results underline the observation that long packets are more likely to be affected by cell contention at the output buffers of the switching and multiplexing equipment. Also, they demonstrate that it is not convenient to send cells at the link bit-rate, as long as packets produce buffer overflow due to the fact that local switching equipment does not usually have large buffers. It is important that ATM drivers should include some facilities to limit the maximum bit rate with which cells are transmitted, or to perform some kind of cell spacing function.

4.3 Extension of experiment 2

The same technique can be used to study the performance of the network access by loading the output port of the local switch. Figure 6 shows the connection set-up of the switch. In this case, point-to-multipoint connections from port 5 are established so that cells re-circulate through the loops and then sent to port 15 in the network. In this way, port 15 is loaded depending on the number of branches directed to it. The counters for the transmitted and dropped cells at port 15 give us the cell loss ratio for the point-to-point reference connection between the local and remote workstations.

Special care must be taken when setting up this experiment. We can use unassigned VCIs at the UNI so that background traffic cells will be dropped at the network switch input port. If current assigned VCIs are used, the network can be overtaken by the excessive amount of traffic and serious problems may arise. Of course, this should not be done without previous agreement from the network provider and the other testbed participants.

5 EXPERIMENT 3. TCP TRAFFIC BETWEEN LOCAL HOSTS

The last experiment intends to measure the throughput of a TCP connection between workstations 1 and 2 when cell losses occur in between.

A simulation study of TCP connections over ATM that share one output link on an ATM switch was presented in Romanov and Floyd (1994). They show the effective throughput for different packet lengths and switch buffer sizes. The effective throughput drops to very low values (about 30%) for small switch buffers and long packets. They also present values for the cell loss rate and packet re-transmission percentage corresponding to TCP connections with different packet lengths and switch buffer sizes. We intend to do this type of study in the experimental ATM network, loading the switch in order to have different levels of congestion, or cell losses, and study the effective throughput, packet losses, and re-transmission percentages for different packet length of the TCP connection. In order to get correct results, we have to adjust the TCP timers as suggested in Romanov and Floyd (1994) and we cannot vary the switch buffer size.

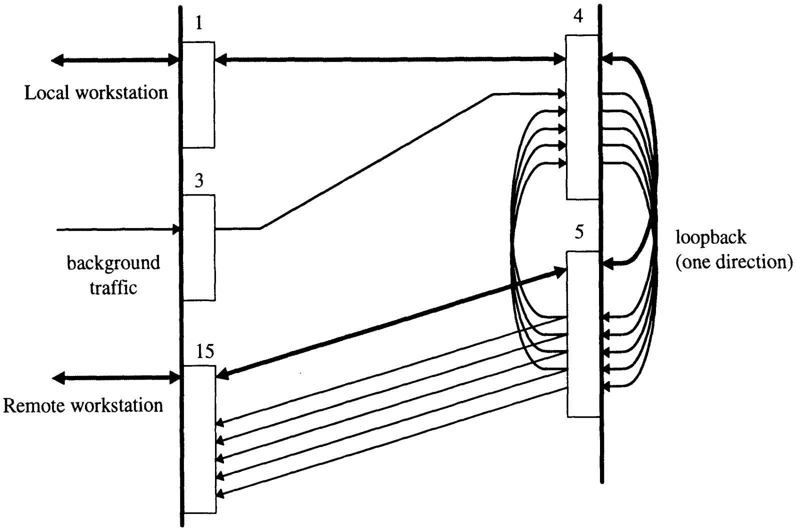


Figure 6 Connection set-up of the ATM switch for experiment 3.

For this experiment a bulk TCP transfer is set up between workstations 1 and 2. In this first case we use standard TCP without modifying the timers and with a buffer of 64 Kbytes and a packet length of 8192 bytes. We experienced that with such long packets the performance was very poor, even though we only measured relatively few cell losses. Only one direction traffic loop was active, so that cell losses could only occur in one direction and only data packets could be lost.

Summarizing some of the results obtained, we measured a throughput of 0.46 Mbps. with a cell loss ratio of $2.4*10^{-5}$ in the receiving buffer and $2.3*10^{-4}$ in the transmitting buffer at port 4. Furthermore, a background traffic of 20 Mbps. was enough to disturb the TCP data transfer. In fact, when a TCP packet of about 172 cells is transmitted at 150 Mbps. and this traffic merges with the background traffic of 20 Mbps cell losses occur. A single cell loss can affect a TCP packet, which will be discarded and re-transmitted.

As we said before, depending on the load in the workstation the background traffic generated varies. Several more measurements gave the results presented in the following table.

Table 4 Throughput between icsibag1 and icsibag2

Background traffic at port 4	*TCP throughput*
11 Mbps.	8.31 Mbps.
20 Mbps.	7.54 Mbps.
22.3 Mbps.	6.88 Mbps.

6 CONCLUSIONS

We presented techniques to measure the performance of a local ATM switch and the ATM cell relay service offered by the carrier to forecast the service performance as the traffic grows. The technique is not expensive as it uses network management tools and software diagnostic applications which are already available in most workstations.

As an alternative to traffic modeling and simulation, traffic emulation can be used to study congestion in ATM gigabit networks. The load of the switch is adjusted by changing the number of re-circulating loops, and by using the available software tools in the workstations, we can measure the actual cell loss ratio for a given PVC. The method allows us to load the switch and emulate the congestion situation causing cell losses and to study the behavior depending on the packet length. For high load we measured cell loss ratio of about 10^{-4} which causes a packet loss of 13% for 2048 byte packets, and for a cell loss ratio of about 10^{-2} a dramatic increase of 90% packet losses was recorded for packets 2048 bytes long.

7 REFERENCES

Atkinson, R. (1994). *Default IP MTU for use over ATM AAL5.* RFC 1626.

ATM Forum (1994). *The ATM Forum ATM User-Network Interface Specification, Ver. 3.1.*

Heinanen, J. (1993). *Multiprotocol Encapsulation over ATM Adaptation Layer 5.* RFC 1483.

ITU-T Recommendation I.363, B-ISDN ATM Adaptation Layer (AAL) Specification, Section 6 (AAL5).

Laubach, M. (1994). *Classical IP and ARP over ATM.* RFC 1577.

Romanov, A. and Floyd, S. (1994). Dynamics of TCP Traffic over ATM Networks, in ACM SIGCOMM, London, September 1994.

Sun Microsystems, (1993a). SunNet Manager 2.2. Programmer's Guide.

Sun Microsystems, (1993b). SunNet Manager 2.2. Reference Guide.

Sun Microsystems, (1993c). SunNet Manager 2.2. User's Guide.

SynOptics, (1994a). ATM Network Management Application.

SynOptics, (1994b). Release Notes for ATM Connection Management System.

SynOptics, (1994c). Using the SBUS ATM Host Interface.

8 BIOGRAPHIES

Jordi Domingo-Pascual is an associate professor of computer science and communications at the Universitat Politècnica de Catalunya (UPC), Barcelona. There, he received the engineering degree in telecommunication (1982) and the Ph. D. Degree in Computer Science (1987). Since 1983 he is working at the Computer Architecture Department. His research topics are Broadband Communications and Applications. Since 1988 he has participated in RACE projects (Technology for ATD and EXPLOIT) and in several Spanish broadband projects (PLANBA). Since 1995 he is working at the Advanced Broadband Communications Center of the University (CABA) which participates in the Spanish National Host and in the PLANBA demonstrator.

Andrés Albanese did his undergraduate studies in electrical engineering at Universidad Central de Venezuela in Caracas. He earned his master's degree in electrical engineering at the University of Texas at Austin (1972), and his Ph. D. degree at Stanford University (1975). He joined Bell Telephone Laboratories (1975-1983). In 1983 he returned to Venezuela for one year to lecture at the Metropolitan University of Caracas. Back to the United States, he joined Bellcore in 1984. In April 1993 he joined the International Computer Science Institute as co-leader of the Networks Group. His credits include over 80 publications and 13 patents in fiber optics, electronics, and networks.

Wieland Holfelder studied Computer Science and Business Administration at the University of Mannheim, Germany (1987-1992). He earned the Master Thesis at the European Networking Center of IBM in Heidelberg, Germany (1993) and the Degree as Diplom Wirtschaftsinformatiker from the University of Mannheim, Germany (1993). During 1994-1995 he had a Ph.D.-Scholarship from the German Academic Exchange Service at the International Computer Science Institute in Berkeley, CA. Since 1995 he has a Research Position at the University of Mannheim and the European Networking Center of IBM in Heidelberg, Germany. His interest and research areas are: Distributed Multimedia Applications, Multimedia Teleconferencing and High-speed Networks.