Project Number:	**RE 1009**
Project Title:	**TEN-34**

Deliverable D11.3

Results of Phase 1 Test Programme

Deliverable Type:	PU - Public
Contractual Date:	30 April 1997
Actual Date:	-
Work Package:	11 - Phase 1 Test Programme
Nature of Deliverable:	RE - Report

Authors:

Michael Behringer, DANTE
Mauro Campanella, INFN Milano
Zlatica Cekro, ULB/STC
Phil Chimento, Univ. Twente, NL
Tiziana Ferrari, INFN/CNAF
Christoph Graf, DANTE
Olav Kvittem, UNINETT
Olivier Martin, CERN
Kevin Meynell, UKERNA
Ramin Najmabadi Kia, ULB/STC
Paulo Neves, FCCN
Victor Reijs, SURFnet

Abstract:

In Deliverable D11.1 a set of ATM experiments was specified as part of the TEN-34 Phase 1 Test Programme over the JAMES ATM network. D11.2 reported on interim results, and this deliverable, D11.3, describes the final results of these experiments.

The main emphasis in this phase lies in examining the underlying technology for its suitability to support advanced applications. Some of the experiments concentrate on fine-tuning systems to maximise performance; others investigate the usability of new technologies and ATM traffic classes.

The results of the experiments have shown that basic ATM services such as CBR are well understood and can be used in a production environment. There is also a better understanding of VBR services and their suitability for IP traffic. The result of our research into more advanced services varied, and in most areas further research is clearly needed. There was significant progress with using signalling for setting up SVCs towards the end of the test phase, and despite the fact that the test environment was not stable enough to be used for production services, it was possible to find a solution to some of the principal problems outlined in D11.2.

With the conclusion of the first phase of the experiments in the TEN-34 project we have proven the basic ATM services to work reliably for IP traffic. We also have gained greater understanding on how to use the more advanced features of ATM. Despite these positive results the overall conclusion of this phase is that advanced ATM features do not work reliably enough yet to be used on a production network.

Keywords:

ATM experiments, IP over ATM, TCP high-speed testing, SVC testing, ARP testing, NHRP testing, ATM Addressing, ATM network management, CDV testing, Native ATM performance testing, IP over VBR testing, RSVP testing

Table of Contents

Executive Summary

The TEN-34 project consists of two parts: The immediate deployment of a high-speed production IP network, and the testing of mostly ATM based advanced network services for future usage on the production network. To avoid interference of potentially unstable experimental services with the production network, these tests are being carried out on a physically separate infrastructure, the JAMES network.

In Deliverable D11.1 a test plan was laid out to precisely define the first set of experiments that needs to be carried out for the development of future services, and D11.2 presented interim results. This deliverable, D11.3, reports the results of these experiments. A brief summary of the results per experiment can be found below.

At the end of the first phase we have obtained a good understanding of the possibilities of ATM services. Basic ATM services are now well understood. The TEN-34 network is already making use of ATM CBR and VBR traffic classes, and there are no problems with those services. VBR VCs are however being used in a very conservative way.

The conclusion for the more advanced services was in most cases that they are not stable enough yet to be run on a production network. We were able to eliminate some of the basic problems which we found for example with SVCs. However, the solutions still rely on idealistic circumstances in the configuration or on sophistication of the upper layer protocols (e.g., TCP). The lack of operational stability remains a serious concern and is holding back deployment of advanced services.

The overlay network provided by JAMES was stable throughout the test programme, but the scope of services is still limited. In addition to a CBR service JAMES offers now also VBR services. Unfortunately this service was not useful for our experiments, as in most countries there is no national VBR service. VBR tests are only useful if the VBR service is end-to-end. Therefore the provisioning of VBR only on the international portion solved only part of the problem, and a full integration of national ATM services is required. SVCs were not available on JAMES at the time of writing this report, a problem which we circumvented by tunnelling the signalling information through the static set of CBR VPs provided by JAMES.

The results of the Phase One Test Programme have shown that more research into the reasons of the unreliability of some services such as SVCs is needed. This will be carried out in the second phase of the Test Programme, along with experiments on other technologies, such as ABR. Generally the second phase of the tests will concentrate on the more advanced ATM features and on ways of providing - albeit limited - additional network services.

1. Summary of Interim Results per Experiment

1. **TCP-UDP performance over ATM**: This experiment confirmed that ATM CBR services are well understood, and the equipment can be used in a way to make optimal use of the available bandwidth. This experiment has concluded.

2. **SVC tunnelling through PVPCs**: The first result here is that switching information can be tunnelled through permanent ATM connections, allowing to treat ATM infrastructure, which is not capable of switching, transparently. The second result is that it is not yet possible to use SVCs in the way it is intended: between end user applications. Usage of SVCs in specific configurations is possible, but not stable enough for an operational service. The reasons for this instability could not be fully investigated. More work will be done in the second phase.

3. **Classical IP and ARP over ATM**: Local tests, which showed no problems in doing address resolution over a LAN, could be confirmed over the wide area VCs. Further investigation into set-ups with more than one server will be conducted in phase two, but no major problems are expected.

4. **IP routing over ATM with NHRP**: The basic operation of the NHRP protocol could be tested, and results with a limited test set-up showed that the protocol works as expected. There were occasional stability problems, which could be due to the usage of SVCs with the related problems mentioned above. This technique is not stable enough yet for an operational service. Further tests with more partners will be conducted in phase two.

5. **European ATM Addressing**: This activity is mainly investigating the addressing plans of NRNs and PNOs in Europe. The basic result is that both addressing schemes, E.164 and NSAP are going to be used in Europe, and that therefore address translation will be necessary to provide ubiquitous service. This has yet to be acknowledged officially by the PNOs. In phase two the main emphasis will be on address translation.

6. **ATM Network Management**: Experiments with SNMP based network management of NRN routers and switches have shown no problems. A network management system was set up to provide users with a full view of the ATM overlay network. These activities will be kept up in phase two. In addition, work is planned to be done on other management platforms such as X.user.

7. **CDV over concatenated ATM networks**: This experiment highlighted that there is an increase in the variation of cell inter-arrival times on a CBR service with each switches passed on the way. There are a number of potential causes, including the switches themselves and differences between PDH and SDH. A possible conclusion could be that on long paths through ATM networks re-shaping must be done occasionally to comply with the traffic contract or alternately, that the traffic descriptors must be 'loose'.

8. **Assessment of ATM/VBR class of service**: Testing of VBR services over JAMES was not possible, as no suitable end-to-end VBR service could be obtained. However, national tests have shown that there is no benefit in using VBR over CBR for IP traffic, or vice versa. If VBR services are used to carry IP traffic, the VBR service should be configured with PCR=SCR and MBS should be as big as possible.

9. **Performance of the Native ATM Protocol**: Due to a lack of native ATM applications, we were not able to test this. The experiment was deferred to phase two.

10. **IP resource reservation over ATM**: Due to a lack of resources no international tests were carried out in this experiment. Local tests in Germany were successful. International tests are planned for the phase two test programme.

11. **ATM Security:** A threat model was developed and the vulnerability of user data flows, signalling flows and management flows was examined. For each of the flows required security services were identified. Further to this theoretical work practical experiments are planned for the second phase of the experiments.

2. Usage of the JAMES Network

The JAMES project provides the basic ATM infrastructure over which the TEN-34 ATM experiments are being carried out. The ATM services that are offered by JAMES are CBR, SMDS, LAN emulation, as of September 1996 an IP service, and as of January 1997 a VBR service. Of these, only CBR and VBR services are of interest to the TEN-34 community (the IP service is not of interest to TEN-34 for testing purposes, as we are interested in the implementation details of IP over ATM, not just in using an IP service). The VBR service could not be used for testing purposes, as the testing of VBR services provides reliable results only if the whole VC is VBR. However, national VBR services were only available in two countries, so that there was only one possible VBR VC end-to-end that could be tested. Unfortunately, in those two countries it was not possible to test due to a limitation of resources on our side. We plan to take up international VBR tests in the next phase.

In addition to those basic services more advanced services would have been needed. The most important ATM service that was not available (apart from the restrictions with VBR) was signalling capabilities of the network. This was not available throughout the first phase of the test programme. The lack of switching capability could to some extent be circumvented by TEN-34 by tunnelling the switching information through the network of CBR VPs. Thus JAMES is used here only as a transmission infrastructure, with all switching being done in the ATM switches of the NRNs. This way we were able to set up an SVC network despite JAMES not being able to do switching directly. The results of these tests are valid nevertheless for the type of equipment used, but it would be desirable to be able to verify the results over a fully switched infrastructure with a diverse range of switches. The JAMES project have assured us that switching is foreseen to become a service over parts of JAMES during the first half of 1997. We plan to conduct more tests in this area, once the service is available.

The concerns about the operational procedures for the JAMES network could not be resolved throughout the first phase of the test programme. The operational procedures for the JAMES network are very basic. There is only one contact person per PNO with no backup specified in case the main contact is not available. There are no service level arrangements such as set-up time for VPs on the JAMES network, so that it is for example not clear how much lead time has to be given for VP delivery. These problems make the planning of an international set of experiments difficult. There were slight improvements in the ease of document handling, but the set-up time for VPs was at times up to two weeks despite the fact that VPs should be available within hours, once the general set-up is agreed. More streamlining of procedures is needed in this area.

To circumvent these problems we requested a large static set of VPs, so that the bandwidth can be allocated to experiments by TEN-34 directly, rather than going through the JAMES procedures in each case. This procedure was generally working and we did get the requested low-speed CBR VPs for this overlay network. There were a few minor problems during the holiday season when contact persons in JAMES were not available and no other responsible person could be found. The lack of a service level agreement with JAMES remains a serious concern, and has negative impact on the JAMES network, because users such as TEN-34 tend to request more bandwidth than they actually need, to be on the safe side. Apart from these problems the operational procedures worked and we did not have major problems in getting the VPs we requested.

It has to be mentioned that the JAMES staff was always helpful and tried to fulfil our requirements to the best of their possibilities. This was also true for non-standard requests, which were dealt with in an unbureaucratic and efficient fashion.

3. Joint Experiments with JAMES

There was good co-operation with JAMES in two areas, in the CDV tests, and in the security experiments. In both of these experiments JAMES representatives participated actively in the experiment. In all other experiments there was no co-operation from JAMES. TEN-34 kept JAMES informed about its plans at all times, and have made several proposals for co-operation. JAMES representatives were invited to all TERENA TF-TEN (the group who carries out the TEN-34 experiments) meetings. Representatives from JAMES were at the TF-TEN meeting on 30-31 October 1996, where significant progress was made in the co-operation, but at no other subsequent TF-TEN meeting.

After this meeting the progress stalled again, and another meeting was held between JAMES and TEN-34 to discuss further co-operation. A set of actions was defined, targeted mainly to resolve the lack of communication from JAMES. The basic idea was for JAMES to provide more input into the joint experiments, and to make their plans more open. All TEN-34 plans, test descriptions and results are publically available on the Internet, and there is an open discussion list, on which JAMES members have been and still are welcome to join. TEN-34 proposed to adapt a similar scheme for JAMES, but up to today there is still no technical information on the planned experiments from JAMES available, nor is TEN-34 informed about experiments that are being carried out within JAMES. Apart from the security activity, all active input and proposals in experiments have come from TEN-34.

A formal co-operation agreement between JAMES and TEN-34 is still being worked on.

4. Conclusions

The work carried out in this framework shows that most of the advanced features of ATM and the new IP protocols are not yet at a state where they can be used safely for operational services. The problem seems to be in most cases that the development of hard- and software is not mature enough. The results of the experiments do however show how to make best use of the existing services (CBR, VBR), and give a good insight into the problems that arise with new technologies.

More work is clearly needed to fully understand the capabilities of ATM networks and of comparable IP services. In some of the areas described above new questions arose during the tests.

There are also a number of technologies which were not yet examined in phase one. Phase two of the project will also investigate other technologies, such as ATM routing and new traffic classes such as ABR. The focus of the tests carried out here is to make experimental services available on the TEN-34 production network. Although the more interesting features of ATM do not yet seem not to be sufficiently stable for an operational service, we will keep on following the developments in ATM and IP related activities. The latest information on our experiments can always be found on the TF-TEN home page (http://www.dante.net/ten-34/tf-ten/).

The following sections detail the results for each of the experiments.

5.1 TCP-UDP/IP Performance over ATM

5.1.1 Experiment Leaders

Mauro Campanella, INFN, Milano
Tiziana Ferrari, INFN/CNAF, Bologna

5.1.2 Summary

Tests gave a straightforward proof of the round trip time impact on the achievable throughput of a one-way TCP/IP connection over an ATM CBR VP. When the RTT of TCP/IP packets is not negligible, the value of the actual maximum window size is the key parameter which guarantees the correct behaviour of the TCP flow control mechanism. The maximum window size should be large enough to allow the sender to generate one packet and receive back the corresponding acknowledgement without stopping the sending process in the meanwhile. But when hosts are connected by long distance ATM VP over JAMES -with round trip times in the range [40..60] msec- the usual window size upper limit (64 Kby) configurable on traditional operating systems is not enough any more. The window scaling option - which permits larger window sizes must be implemented in the operating system. Tests show that with the proper operating system set-up the total bandwidth reserved on the CBR VP on the JAMES infrastructure is available to an application running on top of TCP/IP. In contrast, when hosts have a limited TCP window size the global bandwidth utilisation can increase only if more TCP connections run in parallel.

Second, when the traffic is not on-way, but full-duplex, i.e. it is generated by two data streams in both VP directions, the aggregate throughput increases, but the maximum value measured is still lower than the total amount of bandwidth allocated (i.e. VP_capacity * 2).

Also the bandwidth reservation scheme used for each CBR VC configured on top of the CBR VP is a key issue for performance. In fact, when we deal with Constant Bit Rate VC's, for each of them a static amount of guaranteed bandwidth must be explicitly set. Now, when two or more VC's are enabled to connect concurrently two or more pairs of remote workstations, it's possible to assign the whole channel capacity to each of them, but we can also distribute it to each VC so that globally the sum of the cells/sec assigned to each VC is not greater than the available one. As the tests show, with these two schemes different levels of performance can be measured. Unfortunately results seem to be contradictory, since for each single bandwidth distribution model results change with the traffic pattern generated. This kind of problems, which are very difficult to understand, seem to depend only on the operating systems efficiency and the software of the ATM adapters installed on the hosts.

Finally, as far as UDP/IP is concerned, tests show that for appropriate datagram sizes almost the total available capacity of the VP can be used to successfully transmit UDP datagrams. In the traffic patterns tested the cell drop rate has no impact on the throughput measured for UDP streams.

Finally, during all the tests the ATM service available in the JAMES infrastructure was good, continuous and reliable.

5.1.3 Participants:

INFN (Italy) , UNINETT (Norway), KTH (Sweden) and RedIRIS (Spain).

5.1.4 Dates and phases

The experiment consists of a single phase, divided into two test sessions, each run on a different network topology configuration and by a different set of partners:

test session Italy-Sweden: 15/19 July 1996;
test session Norway-Spain: 22nd July-2nd August 1996.

5.1.5 Goals

The tests have been done to achieve the following targets:

- the monitoring, whenever necessary, of the IP/ATM performance in the JAMES infrastructure through the measurement of the following parameters:

- throughput (data sent/time) for memory-to-memory data transfers over a VP infrastructure with either full bandwidth available or with bandwidth shared by many users;
- IP packets round trip time average and variance;
- CPU utilisation at both the sending and receiving host; packet loss rate.

- the analysis of the network behaviour when the infrastructure is stressed by different traffic patterns. The aspects monitored in the tests sessions, were the following:
  - the fairness of bandwidth distribution when a VP is shared by different applications;
  - the relationship between the average throughput, the peak cell rate on Constant Bit Rate VP:
  - the possible congestion in the switches in the user's and/or JAMES premises.

For each session tests were done generating different patterns of traffic on the VP through JAMES according to the following stream models:

- many-to-one: many hosts sending to a single receiver;
- one-to-many: one host sending to many receivers (to test the bandwidth distribution between each TCP/IP stream);
- one-to-one half and full duplex streams for many TCP connections between the same couple of host (to test the fairness in bandwidth distribution).

- the analysis of the impact of the TCP window-based flow control algorithm on throughput over an ATM VP wide-area connection and on VP with different round trip time.
- the performance comparison of different implementations of the TCP-UDP/IP protocol stack for some operating systems (evaluation of optimised versions).
- the impact of ATM cell loss on throughput when a non reliable datagram protocol (UDP) is used.

5.1.6 Network infrastructure

Test description

The wide area ATM infrastructure operated by JAMES gave the opportunity to analyse the impact of the TCP/IP flow control mechanism on the performance of applications when high-speed links are used. The efficiency of the windowing flow control style was measured by working on the setting of the socket options which directly determine the window size: the send socket buffer size and the receive socket buffer size. Also the impact of the application message size (i.e. of the amount of data written in the kernel memory through a single system call write() on the throughput was measured.

All the tests were done by generating a real data stream between two or more end-points. Different and complex stream topologies were configured in order to stress the switches and to analyse the TCP/IP flow control efficiency.

The public domain benchmarking application Netperf developed at Hewlett Packard was used.

5.1.7 Network configurations

For each test session a different network set-up was configured.

For the experiments between Italy and Sweden, a permanent constant bit rate VP going through Italy, Germany and Sweden was configured on the JAMES side with 24 Mbps of bandwidth capacity (see figure 1).

In contrast, for the test session between Norway and Spain 36,000 cells per second were allocated and on the JAMES side the VP went through Norway, Denmark, Great Britain and Spain as figure 2 shows. The bandwidth of the VPs is stated here either in cells per second or in Mbit/s, depending on the unit that was used by the PNOs.

For the first test session the equipment used on the user local side consisted of one Sparc Station 5 (Solaris 2.5) and one HP 725/75 (HP-UX 9.05) in Sweden; one Sparc Station 20 (Solaris 2.4) and one Silicon Graphics W8C2-1G64 (IRIX 5.3) in Italy.

For the second test session we had two HP9000 (HP-UX 9.05) in Norway and one Sparc Station 20 and Sparc 10 (both with Solaris 2.5) in Spain.

For all the partners participating in the tests FORE ATM equipment (switches ASX-200 and workstation adapters) was used.

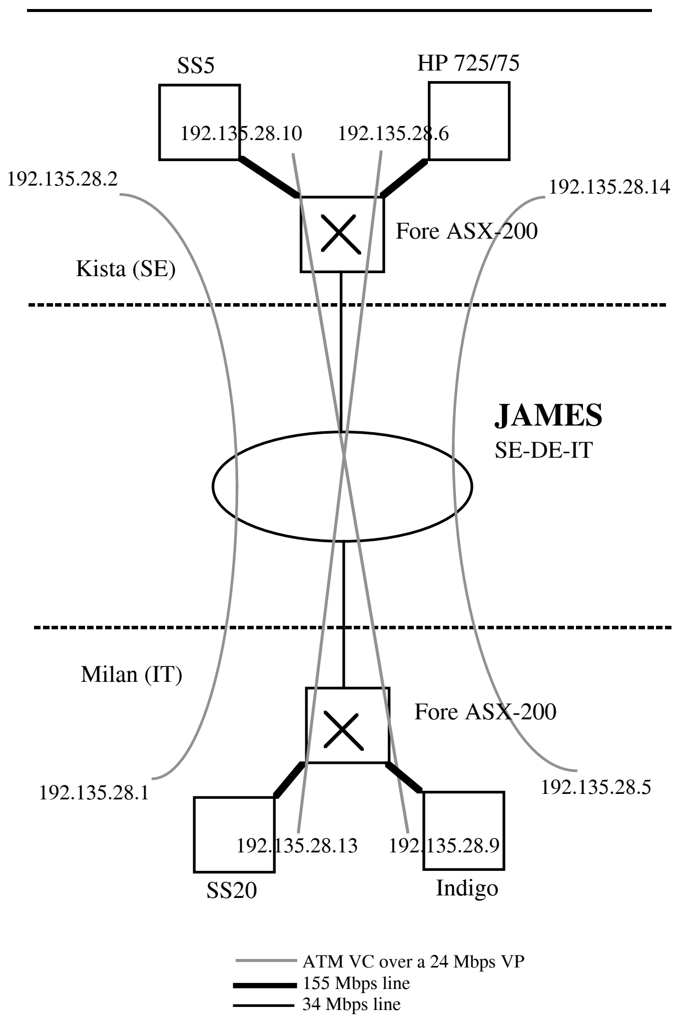


Figure 1: Equipment and network configuration in the test session Italy-Sweden.

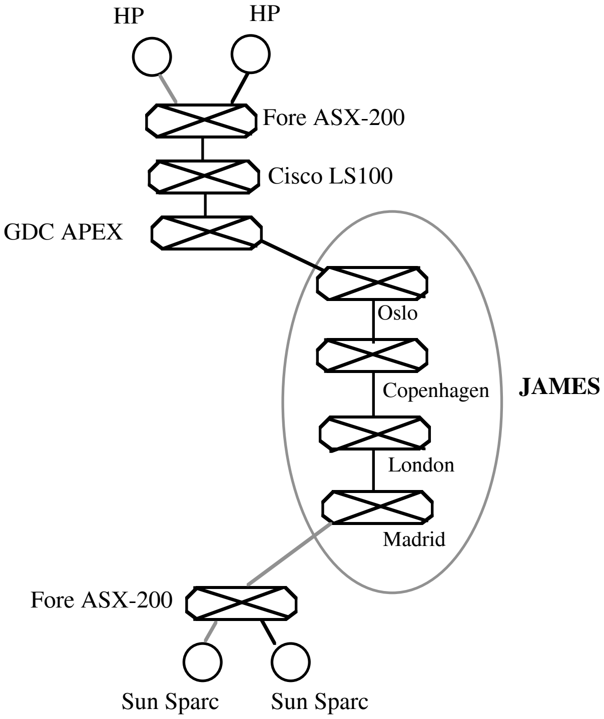


Figure 2: Network configuration in the test session Norway-Spain.

5.1.8 Results and findings

We present in the following the list of some of the most important outcomings of our tests. Important results are divided into two paragraphs.

The first one deals with the tests done with single TCP/IP connections over ATM, while the second illustrates the outcomes when tests have been done with two or more TCP/IP connections drawing a complex topology of connections over the three ATM permanent virtual circuits configured. In the third we analyses the tests done with UDP connections.

1. Traffic performance for a single TCP/IP connection

<u>1.1 Round Trip Time</u>

Measurement of the Round Trip Time (RTT) is a straightforward tool to understand the behaviour of the connections between source and destination. RTT has been traced through the application Ping; of course, in this case, the RTT measured strictly depends on the size of packets generated by ping itself.

The minimum RTT measured in between Italy and Sweden (through Germany, for a total number of two hops inside the public ATM infrastructure) is 40 msec for 10 byte packets, while from Norway to Spain it is 63 msec for 64 byte packets (three hops in-between through Great Britain and Denmark) RTT is linearly dependent on the number of bytes sent down the network. The maximum RTT measured is 96 msec for 61,395 byte packets.

<u>1.2 Maximum throughput and Window Scaling Option</u>

When the RTT of packets is not negligible as in the case of geographical ATM connections through Europe, the size of the Send Socket Buffer Size (ssb size) and of the Receive Socket Buffer Size (rsb size) are the key parameters in order to get the maximum throughput over the ATM pvc.

We can define ssb as the area of the kernel memory in which data are copied as effect of a system call write() generated by the sending application. Symmetrically, the rsb is the area in which data sent to the receiver and coming from the network, are stored. The sizes of ssb and rsb are critical parameters since the Maximum Window Size (MWS), on which the TCP flow control algorithm depends, is a function of ssb and rsb size. For each connection it's calculated by an algorithm according to the operating systems on the hosts. MWS sets the upper bound of the number of TCP/IP packets which can be sent down the network without waiting for the corresponding acknowledgement packet (ack).

Let us suppose that MWS is n byte: if the propagation time of packets is very long compared to the transmission time and the size of the window n, the sender forwards data, but then it stops and waits for ack's backwards. In this case, some available bandwidth is left unused, since during a part of the connection time the sender is idle.

Now, let us give a rough estimate of the lower bound of the window size necessary to prevent the stop-and-wait syndrome.

If for each packet sent an ack is received back -- but this does not apply in any real connection, since the Delayed Ack Algorithm is applied to optimise the mechanism --, the sender can use the whole bandwidth only if after RTT seconds it is still sending data, i.e. the window size Win is at least:

- for the session Norway-Spain (NO-SP):
  Win = (63 msec * 13.824 Mbps) / 8) is approx. 109 Kbyte
- or the session Italy-Sweden (IT-SE):
  Win = (48 msec * 24 Mbps) / 8 is approx 120 Kbyte

Even if the RTT is different in the two test sessions, both Win values are almost the same, since in case 2 the lower propagation time (due to the smaller number of hops involved), is compensated by the higher bandwidth allocated to the VP, which gives a shorter transmission time.

The maximum window size allowed by traditional operating systems is 64 Kbyte (64 Kbyte), which is far less than Win. In order to enlarge the upper bound of the window size, the Window Scaling Option must be implemented in the hosts operating system.

Some of them already include it in the standard version, but some others require a patch or a change of some kernel parameters and the consequent kernel rebuilding.

The relevance of the window scaling option is clear if we compare the test results traced in the two different test sessions. For the tests between Norway and Spain, a patch for Solaris 2.5 has been applied on both platforms and the window scaling option has been enabled also for HP-UX 9.05 on the HP9000's. In this case, thanks to window scaling, 95.5 % of the maximum achievable user throughput was reached. In fact, the cell rate allocated on the VP (namely, 36,000 cps) gives an available bandwidth of 13.824 Mbps on the user level, i.e. without taking into account the cell header. The measured throughput was about 13.2 Mbps and if we take into account the additional overhead due to TCP and IP, we see that almost the total available bandwidth was used.

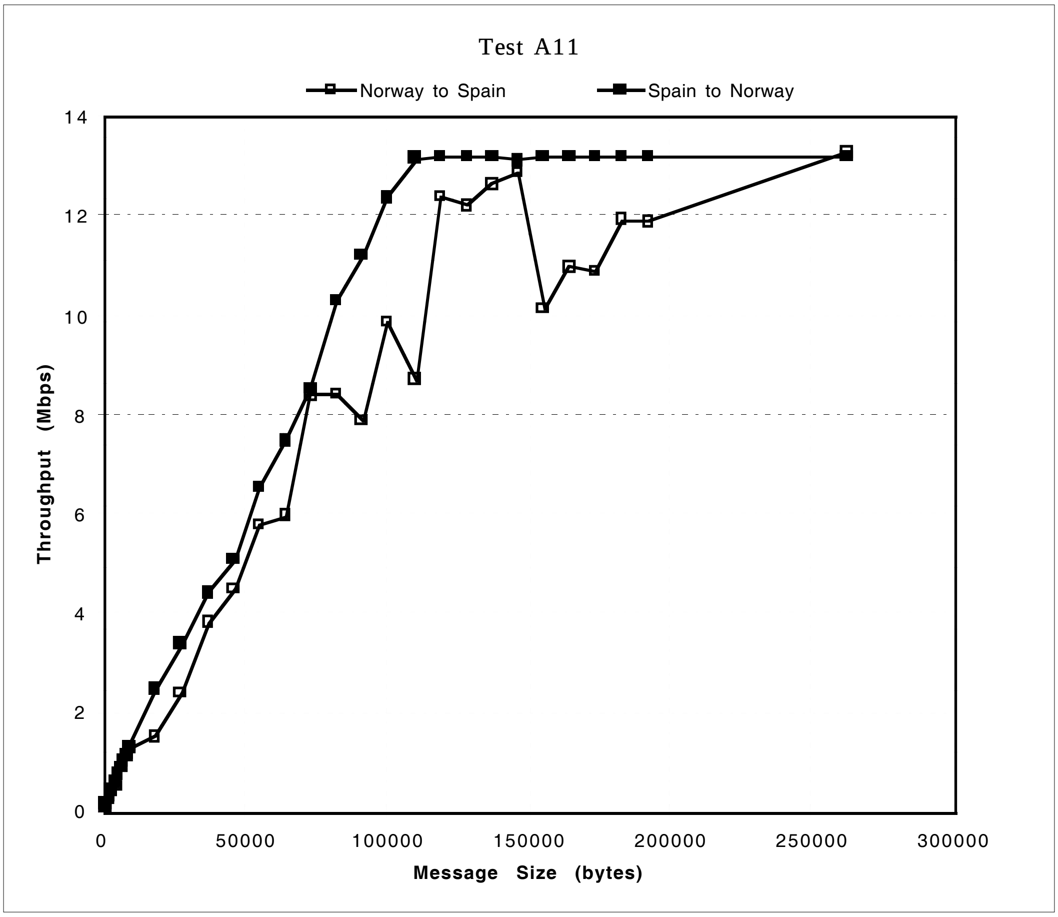


Figure 3: Test session Norway-Spain: Throughput measurement for a one-way TCP connection with variable local ssb/rsb sizes , remote ssb/rsb sizes and mes sage size (ssb = rsb = msg).

As figure 3 shows, when both the ssb and the rsb and the message size are variable with ssb=rsb=message, the maximum is achieved if the parameters sizes are about 120 kbyte, according to our rough estimate of minimum window size Win. Up to that value the throughput increases linearly.

The shape of the function strictly depends on the operating system running on the sending machine: if it runs Solaris 2.5, the throughput increases regularly and after that it is perfectly steady.

In contrast, in the test session IT-SE, the standard versions of Solaris 2.5 and IRIX 5.3 were used. The maximum throughput achieved by one connection was only 8.5 Mbps, which is 35% of the available bandwidth.

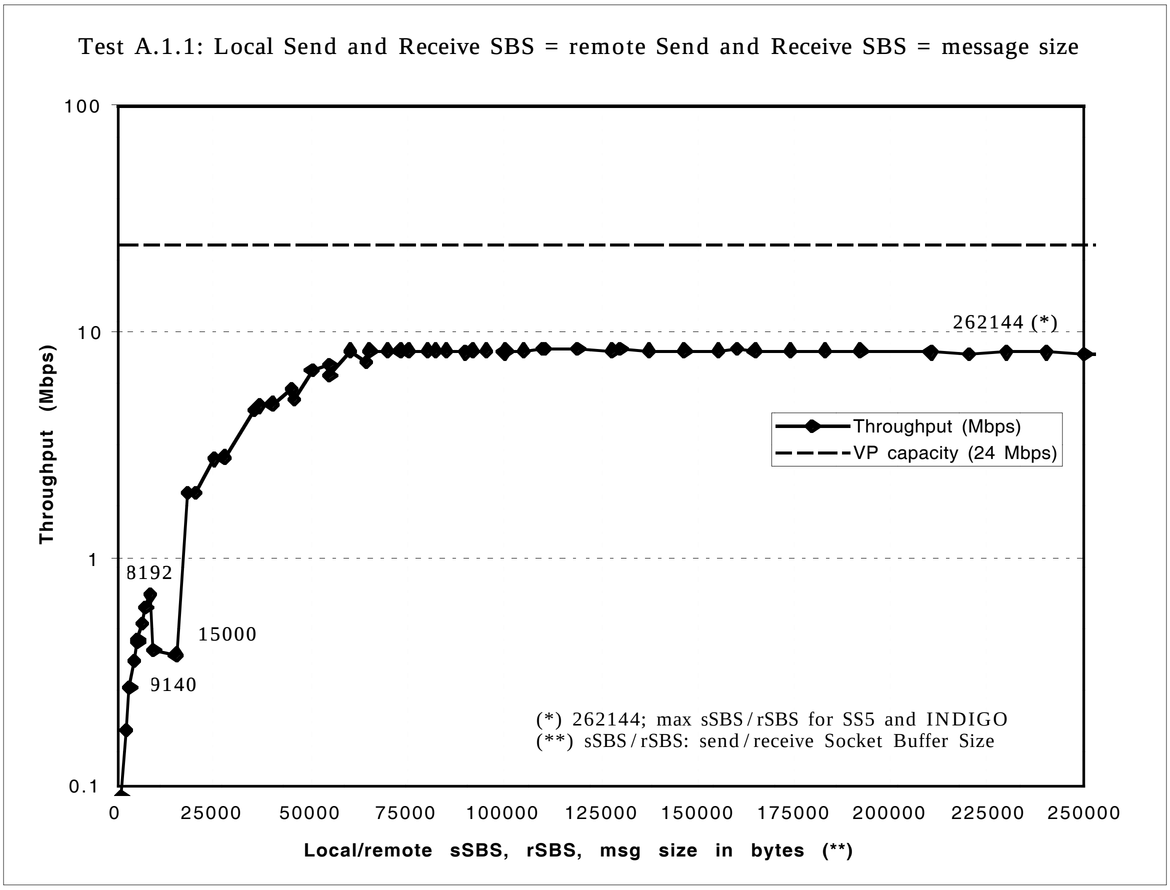


Figure 4: Test session Italy-Sweden: Throughput measurement for a one-way TCP connection with variable local ssb/rsb sizes ,remote ssb/rsb sizes and message size (ssb = rsb = msg).

As figure 4 shows, the throughput increases only for parameter sizes in the range [1..65,000] byte, even if the user application (Netperf) did allow the configuration of both socket and message sizes up to 262,144 byte.

The graph shows clearly that despite of the parameter sizes configured by the user, the operating system did allow only window sizes smaller than 64 Kbyte: for parameters sizes larger than 65,000 byte, the curve shows a constant throughput. Probably some parameters of the operating system like tcp_xmit_hiwat and tcp_recv_hiwat -- both equal to 65,536 byte -- set an upper limit to the window size.

1.3 Send Socket Buffer size and Receive Socket Buffer

The importance of these two parameters strictly depends on the type of operating systems running on the sender and receiver.

For example, from a Sparc Station running Solaris 2.5 to an HP running HP-UX 9.05 (both with window scaling), the throughput is never zero, since for sizes smaller than 64 Kbyte, the throughput is constantly 8.00.

If HP is the sender, the throughput decreases even in the ssb size range [0..65000]. In contrast, if the connection is from an INDIGO with IRIX 5.3 to an SS5 with Solaris 2.5 (not patched), when ssb is constant and rsb vary, the throughput does not change and in this case the only relevant parameter of the connection is ssb.

Therefore, we can say that "the optimal ssb and rsb sizes combination" does not exist, since it only depends on the operating systems and on the algorithms implemented there to set the actual socket buffer sizes as a function of the sizes configured on the application level.
In any case, as we could expect, a symmetrical configuration, i.e.

size(ssb) = size(rsb)

with both sizes configured to the maximum possible value, makes the throughput as high and stable as possible.

This applies to the ssb at the sending side and to the rsb at the receiving side. As far as the ssb and the rsb on the receiver/sender's side are concerned, tests show that the sizes of the rsb on the sending host and of the ssb on the receiving host are irrelevant in the negotiation of the TCP window size. Of course, this result is not a general rule, but it strictly depends on the operating systems present in the testbed.

1.4 No delay

Tests with option No_delay on and off have been done and the corresponding results compared. When the option is on, even small packets can be sent; as a consequence, the Neagles's algorithm -- introduced to make the bandwidth utilisation more efficient and maximise the number of packets with maximum size (MSS), Maximum Segment Size, 9140 byte for ATM) -- is disabled. Tests show that even for small ssb and rsb sizes, this option does not improve the throughput of the connection.

1.5 Message size

With VP bandwidth in the range [0..30] Mbps, the size of the message does not impact the throughput at all. In fact, even with messages smaller than 10 byte, the CPU power of the sending host is still enough to guarantee the maximum throughput.

A small message size makes the application generate an higher number of system calls, that is, more software interrupts and consequently some overhead for their management is added. If the amount of CPU cycles used by the sending process is not high -- this is the case if the VP bandwidth is "low" -- this additional overhead is negligible.

2. Multiple TCP/IP Connections

Meshes of TCP/IP connections were created with different levels of complexity. Four were the types of configurations tested on top of the ATM VP connection:

- 1 bunch of one-way connections between 1 pair of hosts;
- 1 bunch of two-ways connections between 1 pair of hosts;
- 2 bunches of one-way connections between 2 senders and 1 receiver;
- 2 bunches of one-way connections between 1 sender and 2 receivers.

a - Maximum throughput

An increase in the number of connections between hosts (case 1), has the positive effect of increasing the aggregated throughput, i.e. the sum of the all throughputs achieved for each TCP connection.

In test session NO-SP, the throughput reaches 100 % of the available throughput (see figure 5).

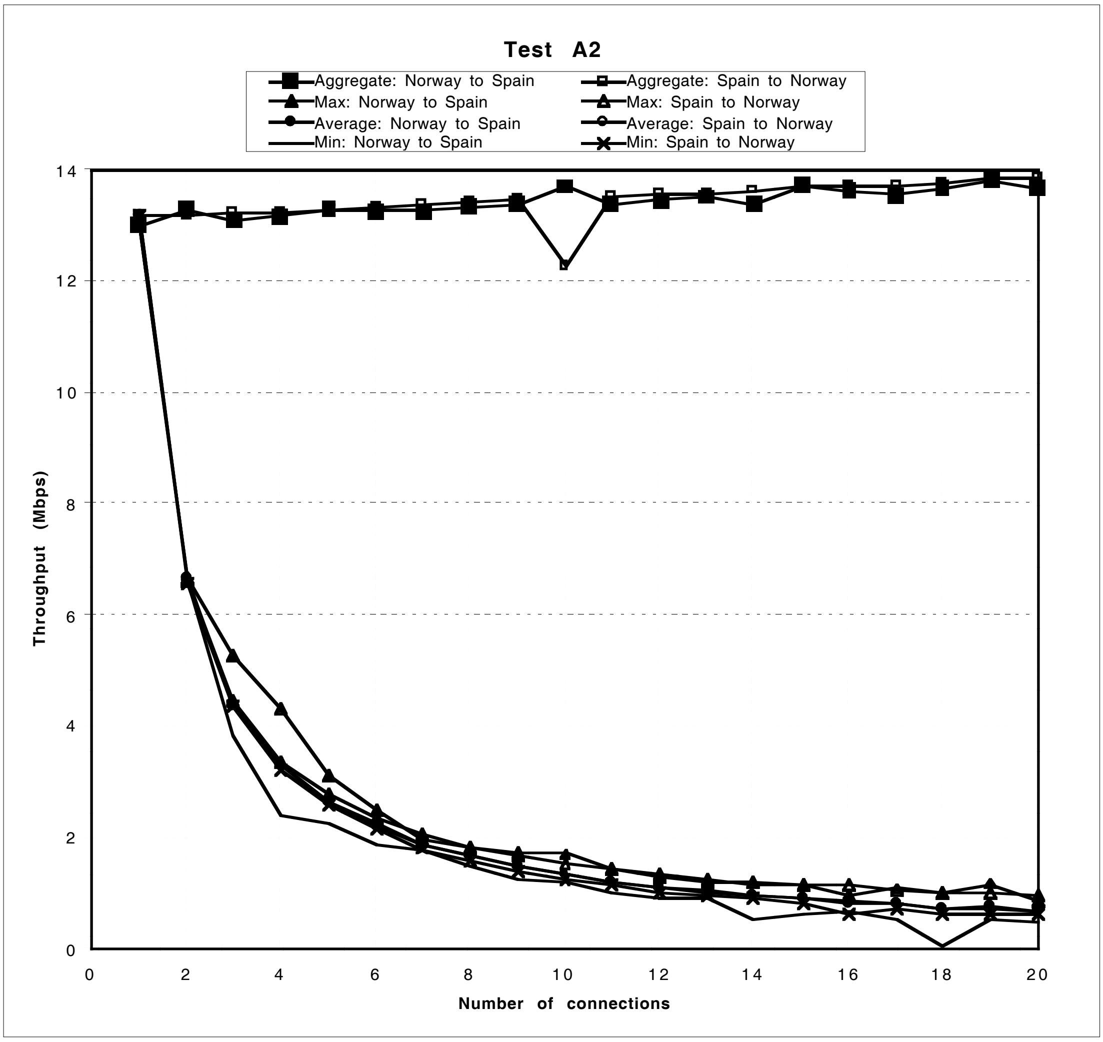


Figure 5: Test session Norway-Spain: Aggregated throughput for a bunch of one-way TCP connections between 1 pair of hosts.

In this case throughput increases slightly when the number of connections goes up to 15. This improvement is even more evident in session IT-SE, as figure 6 shows.

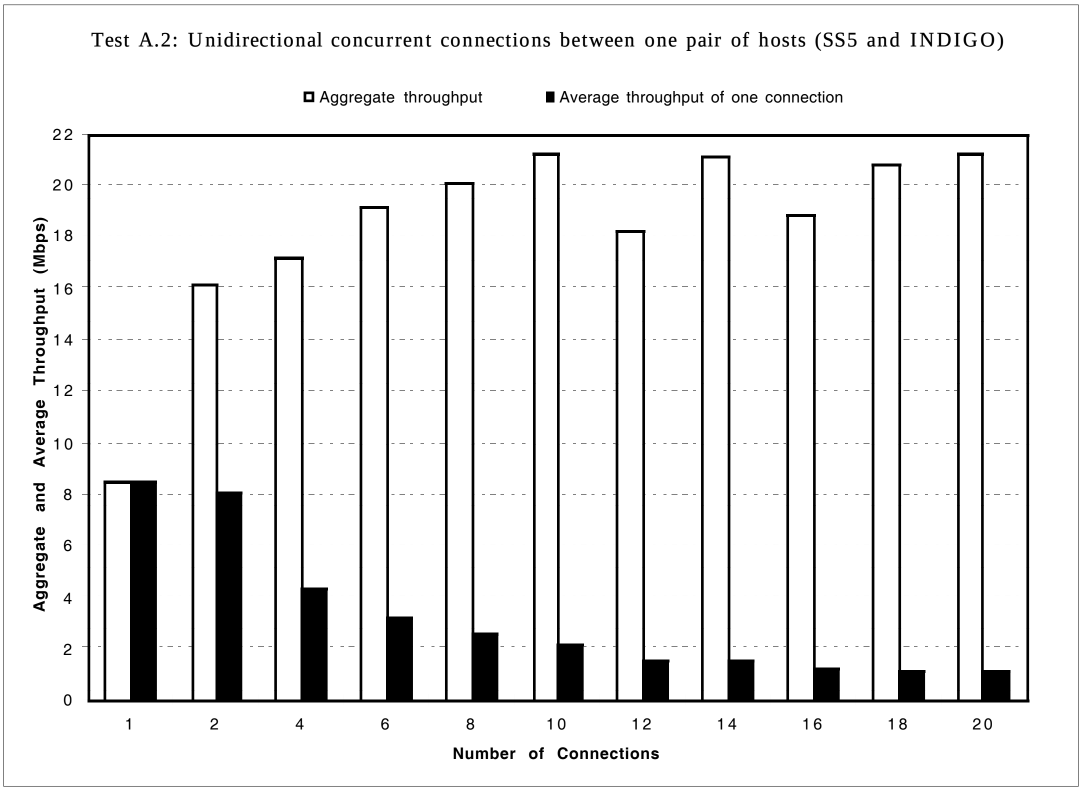


Figure 6: Test session Italy-Sweden: Aggregated throughput for a bunch of one-way TCP connections between 1 pair of hosts.

In this case a single connection is limited to 8.5 Mbps because of the low bound on the maximum window size.

Here the maximum aggregate reaches 21.2 Mbps, which is 88.3 % of the maximum available bandwidth (note that in this case, the aggregate is still much lower than the maximum available). The throughput increases for a number of connections up to 10, after that the aggregate fluctuates around the maximum value erratically. In any case the throughput is fairly distributed among the active TCP connections.

The improvement of the performance with more concurrent TCP connections is a good result, because this model is much more similar to the real Internet traffic patterns, in which typically more users contact 1 or more servers. The big increase measured in session IT-SE can be easily explained: when more concurrent TCP connections are active, the stop-and-wait syndrome on connection i (conn(i)) is statistically compensated by other connections conn(j) whose sender is still sending data to the corresponding receiver.

b - Full duplex bandwidth level of occupancy
When concurrent connections are activated between two hosts in both directions, the aggregated throughput is only about 75% of the maximum achievable throughput, in particular, 35.7 Mbps on the IT-SE VP with 24 Mbps bandwidth in each direction, and 20 Mbps on the NO-SP VP with about 13.8 Mbps, again, in full duplex.

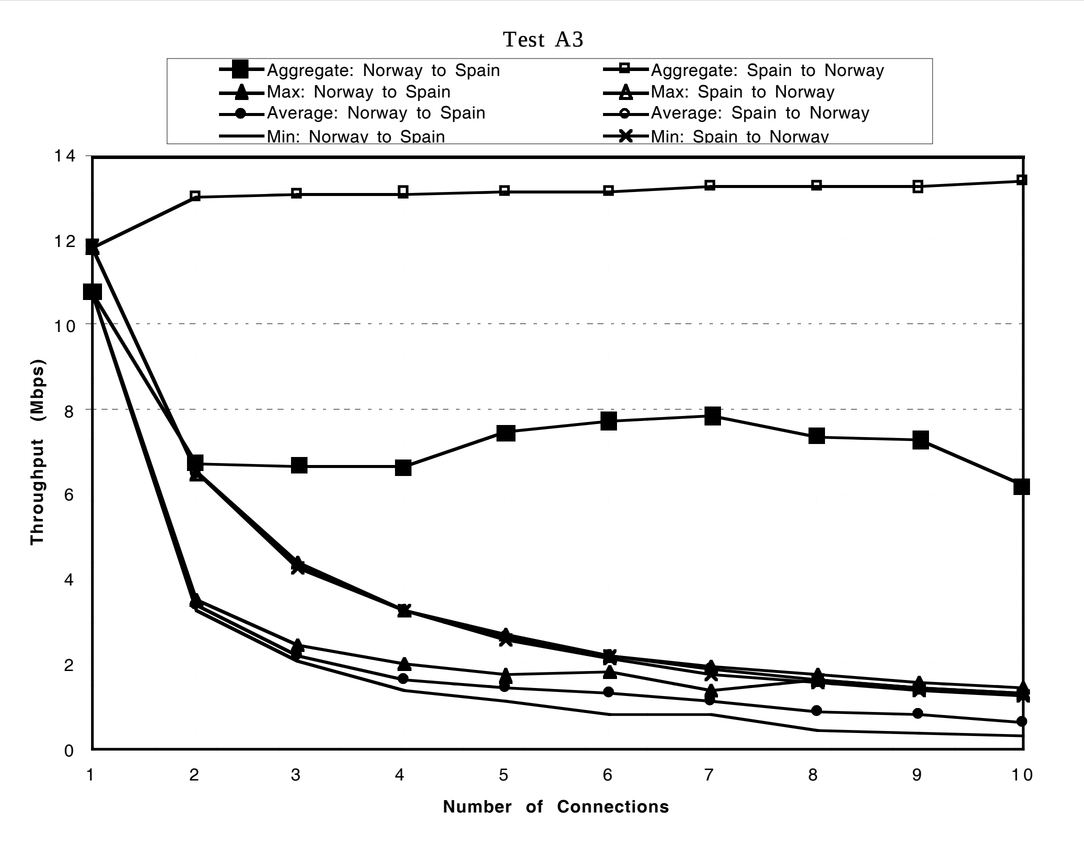


Figure 7: Test session Norway-Spain: Test of aggregate throughput for a bunch of two-way TCP connections between two hosts.

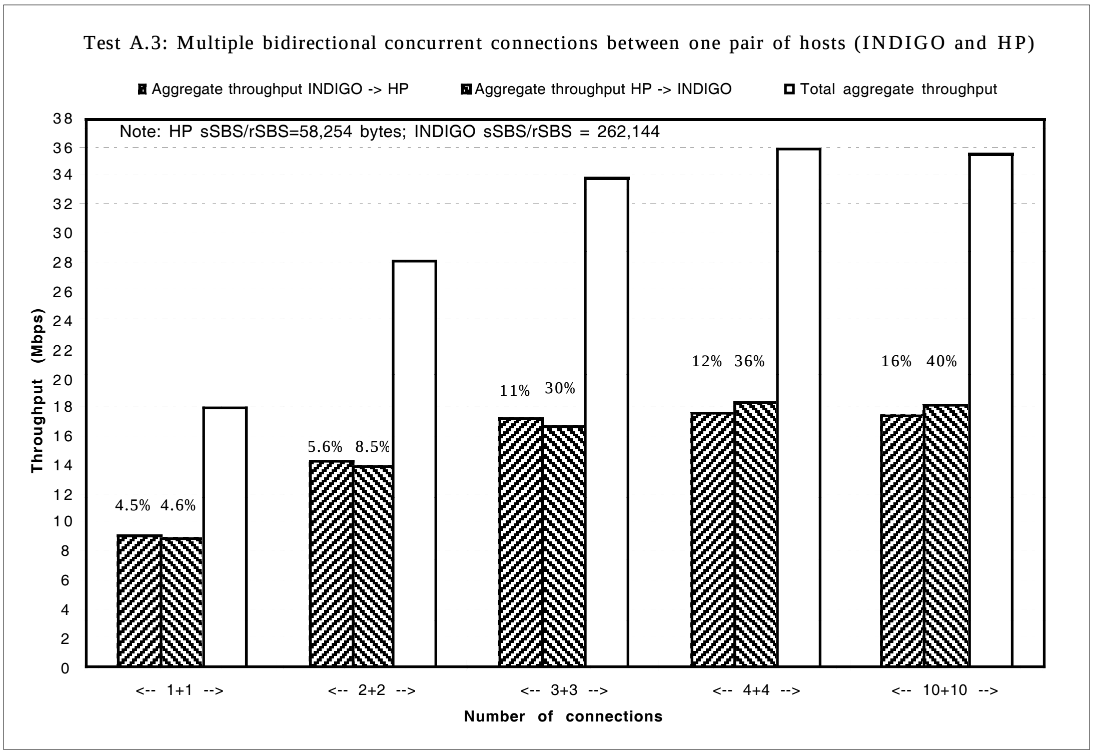


Figure 8: Test session Italy-Sweden: Test of aggregate throughput for a bunch of two-way TCP connections between two hosts.

As figures 7 and 8 show, the aggregate measured for all the one-way connections from a workstation A to a workstation B can be equal or smaller than the equivalent aggregate obtained when only one bunch of one-way connections is run (i.e. we have only half duplex connections). This aggregated value keeps up only for data streams from a Sparc Station 20; in all the other cases it decreases: we lose about 4 Mbps on a 24 Mbps VP and 5 Mbps on the 13 Mbps VP for connections from HP and INDIGO! Even if the aggregate reaches 75% of maximum, we still lose a 1/4 of bandwidth. The reason for this behaviour could not be clarified during the time of the experiments.

c - Peak cell rate configuration on VCC
When according to the configured traffic pattern more than one senders generate data simultaneously to one or more receivers connected by means of ATM VC connections, the right configuration of the VC's and, in particular, the amount of bandwidth assigned to each of them is a key point. For example, let us suppose to have one receiver and two senders which share the same Constant Bit Rate (CBR) VP with bandwidth b.

If we configure two VC's on this VP and we assign b Mbps to each of them, the aggregated throughput of each VC fluctuates and then decreases with the number of running connections, as figure 9 shows.

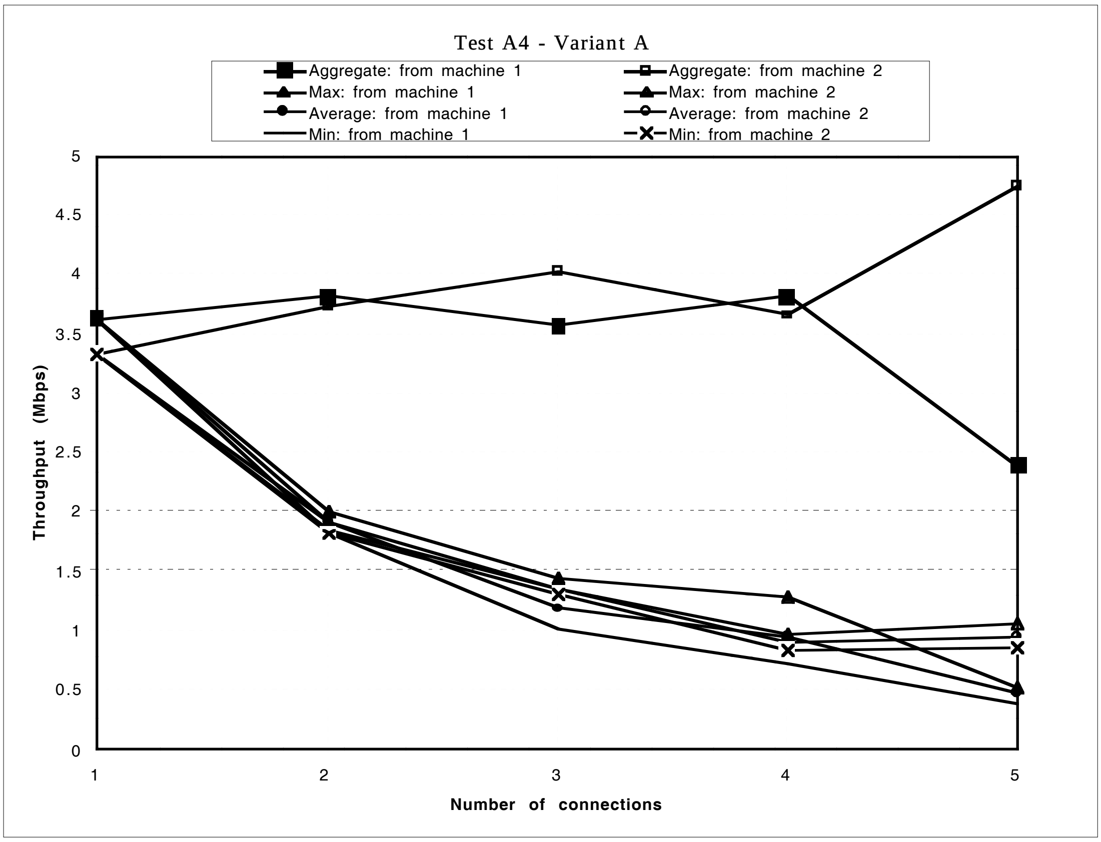


Figure 9: Test session Norway-Spain: aggregate throughput for 2 bunches of TCP connections from 2 senders to 1 receiver and with PCR =b on each VC.

With five concurrent streams on each bunch we lose more than 50% of the available VP bandwidth. In the worse case, e.g.. with a Solaris 2.5 on a SS 5 and HP-UX 9.05 on a HP-9000, all the connections from the HP box are preempted: the number of cells sent on that VC turns to 0 and the bandwidth is only occupied by the SS 5 stream.

In contrast, if only one half of the bandwidth is allocated to each CBR VC, bandwidth is fairly distributed among all the concurrent data flows and the aggregated throughput increases with the number of active streams.

It is interesting to underline that this kind of static bandwidth distribution to n different Constant Bit Rate VC's is highly inefficient when traffic is not equally distributed among the

VC's: if only one sender is active and the other (n-1) are idle, that sender gets only b/n Mbps, where b is the VP bandwidth.

A further remark: peak cell rate, which fixes an upper bound on the rate of the outgoing cells on the hosts, can't be overestimated, i.e. PCR should be according to the following formula:

PCR <= b

Allowing the sender to generate few more cells than what can be allocated, causes an immediate throughput decrease, because of the number of dropped cells, as figure 10 shows.

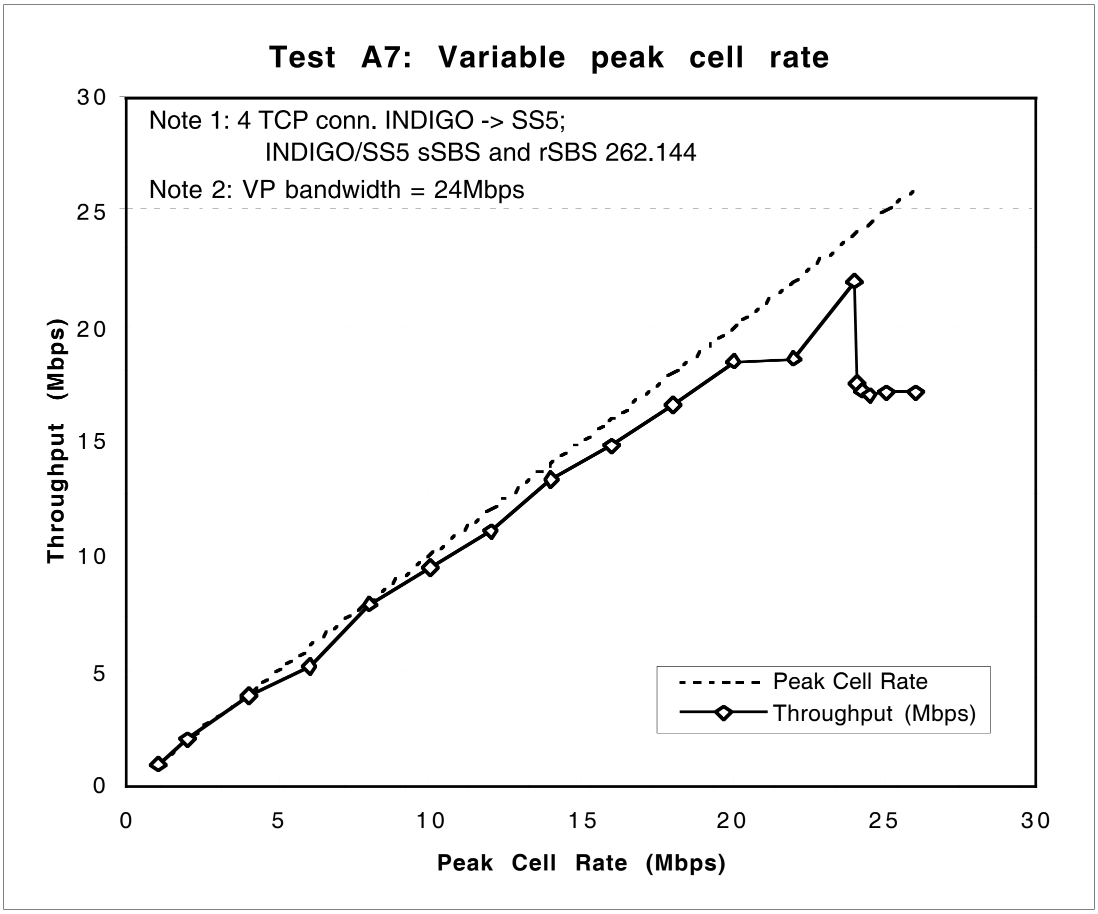


Figure 10: Relationship between throughput achieved by a one-way UDP connection for different datagram sizes and for different peak cell rate values assigned to the VC.

The function shape in the figure also shows that when FORE ATM adapter cards with software versions comparable to the ones present in our testbed, are used, the PCR upper bounds the achievable throughput and the gap between real and theoretical throughput increases with the PCR value.

3. Tests with UDP connections

UDP streams are useful in order to measure the maximum number of received datagrams which are correct, i.e. can be sent down the network without being affected by cell drop, since no flow control algorithms are adopted in this case. The comparison of TCP and UDP tests can show up any limit on the maximum achievable throughput imposed by the TCP flow control mechanism.

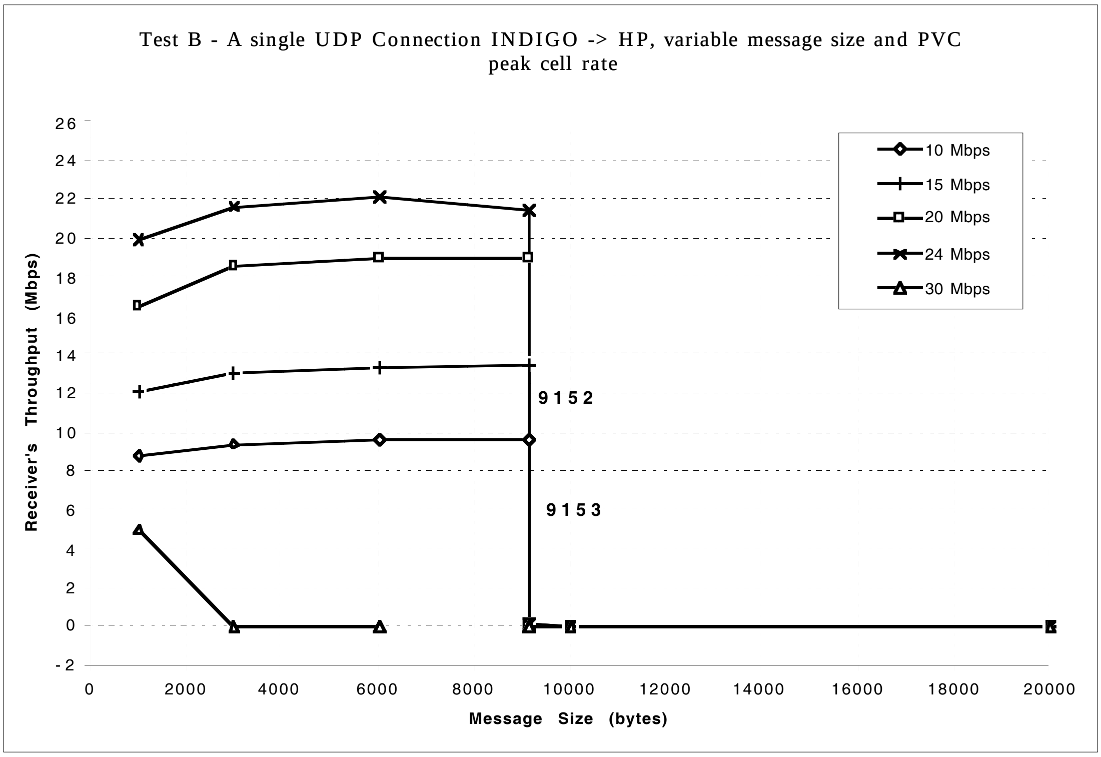


Figure 11: Relationship between variable peak cell rate (PCR) configured for the ATM VC and throughput achieved by a single one-way TCP connection on top of the VC itself.

Figure 11 shows the throughput function shape for different values of Peak Cell Rate (PCR) assigned to the CBR VC when the size of the datagram increases. The interesting range of datagram size is [1..9152] byte.

If the UDP datagram is longer, it does not fit in one ATM MTU (Maximum Transfer Unit) any longer and because of the No_fragment option enabled, the receiver can't assemble the original packet.

In contrast, in the valid range the achieved throughput is still constant and lower than both the available bandwidth and the PCR, independently of the size of the datagram itself. The average number of CPU cycles used by an INDIGO to send the datagrams, is much higher than the one measured for TCP connections: in this case it jumps to 93%.

5.1.9 Relevance for service

The measurement of the TCP and UDP/IP throughput over ATM offers the chance to analyse the network behaviour and the performance level achievable by traditional TCP and UDP applications on a new geographical VP infrastructure under different traffic patterns and equipment configurations.

Through this kind of test it is possible to identify the best network set-up and to find out the whole set of problems due to interoperability problems and to the different levels of efficiency in the ATM equipment of the testbed.

5.1.10 Test related problems and general comments

a - Ping

When ATM pvc are configured, ATM connections con be tested through ping packets. In our multivendor environment, we saw that given three different vendor workstations, let us say wsA, wsB and wsC, if wsA is the sender and the other two are the receivers, ping works only for packet sizes in a fixed limited range.

This range depends both on the sender and on the receiver, since

range (wsA --> wsB) =/= range (wsA --> wsC)

The reason why packets with size larger than a fixed value S do not work is not clear. The cause seems to be dependent only on the end-systems level of interoperability, this for two reasons:

- when the packet size is set to (S+1) byte, the number of outgoing ATM cells is the same as for packets of size S byte.
- the configuration of the PVCs connecting wsA to wsB and C is the same; also the physical path of the cells generated to wsB and wsC is the same.

Pings from the Sun SparcStation 20 did not hang only with packets smaller than connections from/to any remote workstation go into time-out).

b - Cell drop with multiple one-way streams

As explained in the previous paragraph, if more connections are activated between two hosts, the aggregated throughput can increase a lot, but with workstations whose max. window size is limited, it never reaches the available bandwidth and about 12% still is left unused.

The reason of this problems is still not clear, but the monitoring of the cell streams on the ATM interface of the receiver reveals that some cells are regularly dropped by the receiving interface. This could not be fully examined.

c - Throughput loss on two-way connections

The presence of bi-directional connections requires each workstation of the sender/receiver pair to run concurrently both sending and receiving processes, i.e. to manage both software interrupts generated by the system calls of the sending application and hardware interrupts generated by the ATM adapter when IP packets are received. The increased overhead for the interrupt management can explain the increased number of CPU cycles used for this kind of traffic pattern, Anyway, even the increased amount of CPU used (40%) -still below 100%- can't explain why the aggregated throughput of the connections on a single one-way bunch is less than the one measured without the second bunch in the opposite direction. Since on the Sparc Station 20 running Solaris 2.5 is the only platform for which the throughput did not decrease, we could infer that the throughput loss problem is connected with the level of optimisation of the operating system on the sending and receiving machines.

d - Inconsistency of policies for VP bandwidth distribution between VCs

As illustrated in paragraph 4.2 when n senders generate data streams to one receiver, according to the optimum bandwidth allocation scheme, each CBR VC should get b/n Mbps so that throughput on each VC is guaranteed.

The symmetric test, with one sender and n receivers, seems to show the opposite. Let us call b the amount of bandwidth of the CBR VP.

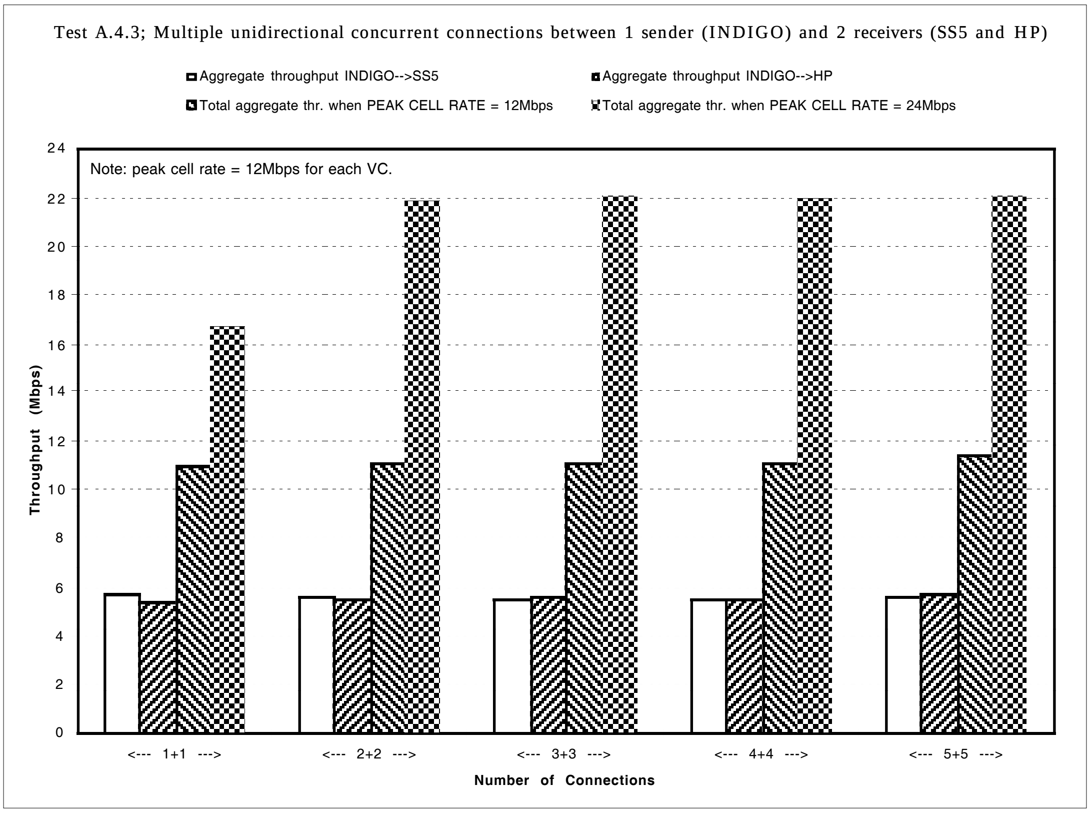


Figure 12: Comparison of aggregated throughputs achieved for 2 distinct bunches of one-way TCP connections from 1 sender to 2 receivers with different peak cell rate configurations.

Figure 12 makes a direct comparison of the two aggregated throughput measurements made either with b or b/n Mbps assigned to each VC (here b=24 Mbps and n=2).

If PCR is b/2, the aggregate does not increase with the number of connections and 50% of the VP bandwidth is left idle. In contrast, if the maximum bandwidth is allocated to all the n VC's, performance improves when the number of streams increases and throughput reaches the usual upper bound measured for a single TCP connection.

More tests in the local area on the user equipment are required for a full understanding.

e - UDP connections with SS 5 running Solaris 2.5

UDP tests run on a SS 5 with Solaris 2.5 show a traffic behaviour different than the one monitored when other platforms are used as senders. First of all, the amount of CPU cycles is comparable to the one measured for TCP (i.e. it is much lower than in the other UDP tests).

Second, all the sent datagrams are received correctly and the outcoming value of bandwidth utilisation is much lower than the available bandwidth.

This only happens when the SS 5 is the sender. If SS 5 is the receiver, CPU utilisation and throughput increase and there are still some datagrams sent which are not received correctly because of cell drop in the network.

For this reason, we can say that UDP tests and the corresponding figures are very dependent on the protocol stack implementation present in the operating system of the sending and receiving hosts.

5.1.11 Further studies

The purpose of this experiment was to figure out the network behaviour under the best possible configuration of traffic patterns and user equipment.

The same test strategy could be deployed to analyse the relationship between the throughput of a TCP-UDP/IP data stream on a long distance ATM VP and the cell drop rate. This could be possible if somewhere on the geographical VP cell drop could be generated on purpose at different rates. Different studies have been done so far in this field, but a TCP-UDP/IP performance over ATM in a degraded network environment test could also clarify the impact of cell drop on high bandwidth*delay VPs.

The same test could be repeated when the Variable Bit Rate service will be available, since up to now, the performance measures done so far depend on the Constant Bit Rate nature of the VP's allocated in the JAMES infrastructure.

Finally, it could be useful to repeat this experiment to figure out also the performance of native ATM applications and to compare it with the results collected for TCP and UDP.

5.1.12 References

[1] Permanent virtual circuits configuration and TCP-UDP/IP performances in a local ATM network;
C.Battista, M.Campanella, T.Ferrari, A.Ghiselli, C.Vistoli.
INFN Internal Note n. 1069, July 1995

[2] Performance evaluation of TCP(UDP)/IP over ATM networks;
S.Dharanikota, K.Maly, C.M.Overstreet, Computer Science Dep., Old Dominion University, Norfolk VA

[3] A Performance Analysis of TCP/IP and UDP/IP Networking Software for the DECstation 5000;
J.Kay, J.Pasquale; Computer Systems Laboratory, Dep. of Computer
Science and Engineering, University of California, San Diego

[4] High Performance TCP in ANSNET;
C.Villamizar, C.Song

[5] High-performance TCP/IP and UDP/IP Networking in DEC OSF/1 for Alpha AXP;
Digital Technical Journal, vol. 5, n. 1, win 1993

[6] How a large ATM MTU causes deadlocks in TCP data transfers;
K.Moldeklev, P.Gunninberg (Norwegian Telecom Research and Swedish Institute of Computer Science).

5.2 SVC Tunnelling through PVPCs

5.2.1 Experiment leader

Christoph Graf, DANTE, Cambridge, UK

5.2.2 Summary of results

It could be shown that the tunnelling of UNI3.0 signalling information across the JAMES network interoperates between all types of switches available to our tests. It can thus be used to bypass the lack of signalling support on the JAMES network and to set-up a SVC network integrating WAN links. The SVC infrastructure set-up in this experiment can be used for subsequent tests in this work package, i.e. ATMARP and NHRP.

The way the IP stack of ATM end systems makes use of the underlying SVC infrastructure is highly limited. Most of our ATM end systems available to SVC tests request best effort SVCs of traffic class UBR only, without any flow and congestion control. All our switches are able to handle UBR requests only. This works fine in uncongested LANs, but is problematic for operation across policed CBR and VBR WAN links as the end system will almost certainly violate the VP contract resulting in severe cell loss. As a result, the SVC network can only be used to carry IP traffic of low bandwidth using small packets.

Our tests show that per-VP traffic shaping on the switch connected to the policed WAN link can be used to shape "well behaved" UBR traffic flows into CBR WAN VPs. TCP with its intrinsic flow and congestion control falls into this category. Once available, the use of ABR SVCs instead of UBR SVCs, together with reshaping of the cell stream on the switch connected to the PNO will allow for all types of traffic flows to get a fair share of the available bandwidth.

Deployment of SVCs in a production environment is strongly discouraged as reliability problems and questions concerning too long set-up times remain unresolved and should be studied further.

5.2.3 Participants to the experiment

- ACONET (AT)
  - Gerald Hanusch, Universitaet Linz
  - Guenther Schmittner, Universitaet Linz
- BELGACOM (BE2)
  - Jan van Ruymbeke, Belgacom
- DFN (DE)
  - Robert Stoy, RUS
- INFN (IT)
  - Mauro Campanella, INFN
  - Diego Colombo, INFN
  - Tiziana Ferrari, INFN/CNAF
  - Simone Maggi, INFN
  - Stefania Alborghetti, INFN
- RCCN (PT)
  - JosÈ Vilela, RCCN
  - Paulo Neves, RCCN
- REDIRIS (ES)
  - Celestino Tomas, REDIRIS
- RESTENA (LU)
  - Alain Frieden, RESTENA
- SWITCH (CH)
  - Simon Leinen, SWITCH
- UKERNA (UK)
  - Christoph Graf, DANTE
- ULB (BE)
  - Ramin Najmabadi, ULB
- UNINETT (NO)
  - Olaf Kvittem, UNINETT
  - Vegard Engen, UNINETT

5.2.4 Dates and phases

Phase one: Set-up and test of local SVC infrastructure
Date: August 96 - March 97 (individual to each participating site)
Duration: approx. 3 weeks

Phase two: Pairwise interconnection of participants over JAMES
Date: August 96 - March 97 (individual to each pair of participating sites)
Duration: approx. 2 weeks

Phase three: Full interconnection of all participants over JAMES
Date: Mid-November 96 - March 97

5.2.5 Network infrastructure

None in phase one.

The second and third phase require VPs (CBR or VBR) of 2 Mbps to interconnect the participants pairwise. The following VPs are or were used in the experiment (not necessarily exclusively):

VP	start date	end date	SVC up
AT-BE2	21/02/97	28/02/97	24/02/97
AT-CH	21/02/97	31/03/97	21/02/97
AT-DE	19/08/96	31/03/97	29/08/96
AT-IT	20/09/96	31/03/97	03/10/96
BE-DE	09/09/96	31/03/97	08/10/96
BE2-DE	24/02/97	28/02/97	24/02/97
CH-IT	05/11/96	31/03/97	06/11/97
DE-LU	01/10/96	31/03/97	25/10/96
DE-NO	28/01/97	31/03/97	28/01/97
ES-PT	27/02/97	31/03/97	28/02/97
ES-UK	27/01/97	31/03/97	27/01/97
NO-UK	24/01/97	31/03/97	24/01/97

5.2.6 Results and findings

Map of SVC connected sites in phase 3

```
PT           BE  BE2   ______
 |           \   / \  /      \
ES---UK---NO---DE---AT---IT---CH
                |
               LU
```

Set-up of ATM equipment in all sites in phase 3

Measurements in phase 1 and 2 are based on slightly different configurations, as indicated in the relevant paragraphs below.

<u>ACONET (AT)</u>

```
JAMES--(STM-1)--GDC APEX--(STM-1)--LS1010
                                     |
                                  (STM-1)
                                     |
             Signalling endpoint --->|
                                  LS1010--(STM-1)--Cisco7010 "jkurtl99"
                                     |
                                  (STM-1)
                                     |
                                   LS100---(STM-1)--Linux "alijku65"
```

- alijku65: 193.246.0.20 Pentium 133MHz, Linux 2.0.25 ENI-155P-MF1 media=oc3 multimode ATM for Linux 0.26
- jkurtl99: 193.246.0.22 Cisco 7010 RP/SSP with AIP interface at 155 Mbit/s IOS (tm) 7000 Software (C7000-JS-M), Version 11.2(4.4)F AIP, hw 1.2, board rev. C0, sw 10.18
- GDC APEX-Mac (connected to JAMES at 155 Mbit/s (STM-1 port 1/0) Software version 4.3.0-A11/Rev E
- Cisco Lightstream 1010 (connected to GDC) IOS (tm) PNNI Software (LS1010-WP-M), Version 11.1(8) RELEASE SOFTWARE (fc1) ASP, hw 3.2, FeatureCard1
- Cisco Lightstream 1010 (connected to 1010) IOS (tm) PNNI Software (LS1010-WP-M), Version 11.2(2)WA3(1a) RELEASE SOFTWARE ASP, hw 3.2, FeatureCard1
- Cisco Lightstream 100 (connected to 1010) LS100 Software Version 3.1(2)

BELGACOM (BE2)

```
JAMES---(STM-1)---ASX200---(STM-1)---SUN
```

- Fore ASX200
- Sun Sparcstation

DFN/RUS (DE)

```
              "ksatm3"        "tencisco1"
JAMES--(STM1)--LS1010--(STM1)--Cisco7000
                  |
                  `----(STM1)--Sun
                              "tensun1"
```

- tensun1: 193.246.0.54 Sun Sparcstation 2, SunSolaris 2.5.1 FORE sba-200e media=oc3 hw=0.2.0 fw=3.0.0, oc3rev=48 ForeThought_3.0.1b (1.28)
- tencisco1: 193.246.0.55 Cisco 7000 with AIP interface at 155 Mbit/s IOS (tm) GS Software (GS7-J-M), Version 11.1(8), RELEASE SOFTWARE (fc1)
- ksatm3: CISCO LS1010, IOS (tm) PNNI Software (LS1010-WP-M), Version 11.1(8), RELEASE SOFTWARE (fc1)

INFN (IT)

```
                         "cisc75misvc"
JAMES--(E3)--LS1010--(E3)--Cisco7507
              |
              `---(STM-1)--ASX20
                             | |
                             | `------(STM-1)--Sun "sunatmsvc"
                             |
                             `--------(STM-1)--Sgi "sgimidasvc"
```

- sunatmsvc: 193.246.0.129 Sun SparcStation 20, SunSolaris 2.4 FORE sba-200e media=oc3 multimode A_ForeThought_3.0.1(1.28)
- cisc75misvc: 193.246.0.132 Cisco 7507 with AIP interface at 34 Mbit/s IOS (tm) GS Software (RSP-J-M), Version 11.1(8), RELEASE SOFTWARE (fc1) AIP, hw 1.3, sw 20.09
- sgimidasvc:193.246.0.130 SGI Indy, IRIX 5.3, GIA-200 adapter, 155Mbit/s
- Fore Asx-200, Hardware version 1.0, Software version ForeThought_3.4.0 (1.29)
- Cisco Lightstream 1010, IOS (tm) PNNI Software (LS1010-WP-M), Version 11.1(8) RELEASE SOFTWARE (fc1) ASP,hw 3.2

RCCN (PT)

```
JAMES--(E3)--ASX200--(STM-1)--Sun "deimos"
```

- deimos: 193.246.0.73, Sun SparcStation 20, SunSolaris 2.5, FORE sba-200e media=oc3 multimode A_ForeThought_4.0.2 (1.26)
- Fore Asx-200BX Hardware version 1.0, Software version S_ForeThought_4.0.2 (1.15)

REDIRIS (ES)

```
JAMES--(STM-1)--ASX200--(STM-1)--Sun
```

- SparcStation
- Fore asx200bx, S_ForeThought_4.0.1 (1.23)

RESTENA (LU)

```
    JAMES--(E3)--ASX200--(STM-1)--Sun
```

- Sparc20 running Solaris 2.5 Fore SBA-100/-200 ATM SBus Adapter running A_ForeThought_4.0.0 (1.30)
- FORE ASX200BX S_ForeThought_4.0.0 (1.30) (asx200bx) (ATM SWITCH)

SWITCH (CH)

```
                   "castor"        "popocatepetl"
    JAMES--(STM-1)--LS1010--(STM-1)--Cisco7505
                      |
                      |
             "netmon" `----(STM-1)--Sun
```

- popocatepetl.svc.tf-ten.switch.ch: 193.246.0.81 Cisco 7505 IOS 11.2(4.1)
- netmon.svc.tf-ten.switch.ch: 193.246.0.82 Sun ULTRAstation 1/170 Solaris 2.5.1 SunATM 2.0
- castor.svc.tf-ten.switch.ch: 193.246.0.83 Cisco LS1010 IOS 11.1(8) PNNI

UKERNA (UK)

```
                      "lemon"   "coney"
    JAMES---(STM-1)---ASX200BX---SUN
```

- lemon: 193.246.0.226, Fore asx200bx, Hardware version 1.0, Software version S_ForeThought_4.0.1 (1.5)
- coney: Sparcstation 5, Solaris 2.4, sba-200e media=oc3 hw=0.2.0 fw=3.0.0 serial=8177 oc3rev=48 slot=1, ForeThought_3.0.2b (1.12)

ULB/STC (BE)

```
    JAMES---(E3)---LS100---(STM-1)---SUN
```

- Sun SparcStation LX, Solaris 2.5, Fore SBA-200 155Mbit/s
- Cisco LightStream 100, version 3.1(2)

UNINETT (NO)

```
    JAMES--E3--GDC--(STM-1)--LS100--(STM-1)--Cisco7000
                |              |
                |              `---(STM-1)--HP(azur)
                |
                `--(E3)--GDC--(STM-1)--LS100
                                        | |
                                        | `--(STM-1)--cisco(trd-gw5)
                                        |
                                        `-----(STM-1)--HP(lunde)
```

- lunde: 193.246.0.178 HP HP-9000/735 FORE/HP card A_ForeThought_3.0
- azur: HP HP-9000/735 FORE/HP card A_ForeThought_3.0
- osloS-gw : cisco 7000 with AIP and sw 11.1(9)
- trd-gw5 : 193.246.0.177 cisco 4700
- GDC APEX - Research ATM backbone by Telenor
- LS100 - Cisco Lightstream 100 Software Version 3.1(1)

Common set-up properties of IP hosts

One IP interface on each workstation, router and most switches was configured with an address from one single LIS (logical IP subnetwork). No IP routing protocol is needed in this case, as all systems will set up direct connections to all partner sites. Classical IP with static ARP tables was used to allow all hosts to interconnect with AAL5 based UBR SVCs across our SVC network. It is, however, required, that the ATM address of each host in the LIS be known on each host. Static IP to ATM address (NSAP) mappings were used in this experiment. The ATM addresses of all hosts were collected, published on a web page and subsequently configured on all IP hosts (see annex).

Common set-up properties of switches

In order to establish SVCs between ATM end systems, the switches involved must know the path towards the destination. This is the task of the routing protocol. IISP (interim interswitch signalling protocol) was used in our experiment. It is based on static configuration of NSAP prefixes manually entered on all switches. One NSAP prefix covering all systems in a given site was collected and distributed in the same way as the IP to NSAP mapping.

No traffic shaping was performed on the switches in phase one and two. Where possible, the switches were reconfigured for phase three tests to perform per VP traffic shaping on the interface towards JAMES. Per interface shaping was used in those cases where per VP shaping was unavailable.

Phase 1: results of local SVC tests

All participating sites were able to establish SVC service between their local ATM end systems.

The following average set-up times were measured at INFN. The ATM end systems were connected by a single intermediate switch. Since the FORE switch is equipped with an IP interface too, it was included in these set-up tests. The set-up times were measured using the standard UNIX tool 'ping'. After ping sends the first IP packet, a SVC is established and then the packet is sent along that path. The round trip time (RTT) of the first packet comprises thus the SVC set-up time plus the RTT of the IP packet along the SVC. Subsequent packets will use the established path and do not need to set up a SVC. By subtracting the RTT of subsequent packets, the SVC set-up time can be obtained:

Avg. times (ms)	SUN-SGI	SUN-CISCO	SUN-SWITCH	SGI-SUN	SGI-CISCO	SGI-SWITCH
SVC set-up	15.97	18.53	18.6	16.73	18.95	19.88
Next RTT	1	1	1	1	1	1

All local measurements in other sites follow the pattern above, except for two hosts: The Linux system at Linz yields somewhat higher times, while tensun4 (with ENI ATM adapter) in Stuttgart needs about 1100 ms to establish SVCs with other local hosts. The reason is not known, but it is expected to be caused by inefficient driver software. Tensun4 was subsequently removed from the SVC network. Cisco routers show another anomaly: they usually discard the first packet, when a SVC has to be established to transport it.

Phase 2: results of SVC tests crossing one WAN link

The following average set-up times were measured at INFN. They refer to SVCs set up between the SUN or SGI workstation at INFN and the Linux workstation or Cisco router in Linz available for our tests. Since the SVC tunnel ends were on the Fore switch in Italy and on the LS100 switch in Austria, two switches were involved in handling the SVC set-up. The time measurements were obtained again in the same way as in phase one testing:

Avg. times (ms)	SUN-Linux(AT)	SUN-CISCO(AT)	SGI-Linux(AT)	SGI-CISCO(AT)
SVC set-up	219.314	96.112	217.117	102.007
Next RTT Avg.	17	17	18	19

We were able to set-up SVC tunnelling on all links mentioned above. Again, the increase in set-up time including the Linux system in AT can be observed. Communication over SVC was possible between all involved types of end systems available to our experiment.

The usefulness of those SVCs is strictly limited to low bandwidth applications using only small packet sizes. The reason is, that the hosts request and get best effort UBR SVCs. Intermediate switches always grant such requests, regardless of the available bandwidth, as UBR VCs do not require the reservation of bandwidth. Depending on the policing policy (CDVT/BT) applied to JAMES VPs and characteristics of the sending host (performance, physical medium, driver implementation, etc.) packets exceeding some size will be lost due to policing. This limit varies heavily between about 120 bytes on the link BE-DE and no noticeable limit on the link AT-IT.

Phase 3: results of SVC tests crossing multiple WAN links

Unlike the tests in phase one and two, the switches were reconfigured for phase three to shape outgoing cells to JAMES to conform with the VP contract, where possible. The limitation of IP packetsize disappears with this reconfiguration for almost all connections on our SVC network. Per VP shaping is being used on Fore switches (ES, LU, PT, UK), while, in absence of this possibility, per interface shaping had to be chosen on Cisco LS1010 switches (AT, CH, DE, IT). No shaping support is available on cisco LS100 switches and older versions of Fore interface cards (BE, BE2, NO). Details about the phase three experiment:

- Tool used for testing connectivity and measuring set-up times: standard unix tool "ping" with default packet size of 64 bytes ICMP payload.
- Connectivity was considered established, if at least 2 out of 5 ping packets were returned from the remote host within 10 seconds from the last packet submission.
- If not already established, the first packet of a ping sequence opens prior to its submission a SVC to the remote host, while subsequent packets will make use of this connection without the need to establish a SVC. Thus, the delay difference in the round trip time (RTT) between the slowest and the fastest packet of such a ping sequence is a good estimate of the SVC set-up time and the results below use this calculation method.
- Test duration was 2hrs, average sample size for all results is about 20 values.
- Not all links and configurations were fully operational during the test period, thus BE, BE2, ES and PT could not or not reliably be reached.
- The tests were carried out from systems in CH, DE, IT, LU and UK.
- The results given below are median values.
- Empty fields indicate that no connection could be established in the desired direction.

local area connection	remote connection	same host connection

Colour legend

193.246.0.xx From: To:	DE (.54)	CH (.82)	IT (.129)	IT (.130)	LU (.145)	UK (.225)
AT (.20)	296	1521	219	346	431	
AT (.22)		44	68	65	150	324
DE (.54)	4	70	98	102	99	282
DE (.55)	17	75	101	105	99	282
CH (.81)			48	47	174	356
CH (.82)	168		37	41	146	377
CH (.83)		9	43	44		345
ES (.100)	232	377			471	
ES (.101)						125
ES (.102)	1287			1355		1116
IT (.129)	90	47	29	28	183	371
IT (.130)	88	46	27		171	360
IT (.132)		25	31	29	160	
IT (.133)				28		416
IT (.134)	100	47	29	28	183	371
LU (.144)					37	
LU (.145)	98	172	171	186		
NO (.177)		302	262	302	569	226
NO (.178)	369		318	331		
NO (.184)	196	250	280	285	284	232
UK (.225)	271	573	351	369	336	
UK (.226)	266	406	360	381	428	26

Table 2.1: SVC Set-up Times in ms (unloaded network)

193.246.0.xx From: To:	DE (.54)	CH (.82)	IT (.129)	IT (.130)	LU (.145)	UK (.225)
AT (.20)	38	12	18	26	58	
AT (.22)		12	18	25	59	122
DE (.54)	1	47	52	54	24	87
DE (.55)	2	47	52	53	24	87
CH (.81)			8	14	80	143
CH (.82)	47		7	13	79	142
CH (.83)		1	8	13		143
ES (.100)	130	174			151	
ES (.101)	130	174				44
ES (.102)	130			190		43
IT (.129)	53	7		2	74	137
IT (.130)	53	7	1		74	
IT (.132)		7	1	3	74	
IT (.133)				4		137
IT (.134)	53	7	1	3	74	138
LU (.144)					2	
LU (.145)	25	68	74	77		109
NO (.177)		90	96	99	68	54
NO (.178)	47		96	101		
NO (.184)	41	84	90	93	62	48
UK (.225)	88	142	137	141	109	
UK (.226)	89	132	137	144	110	2

Table 2.2: Round Trip Times in ms (SVC already established)

193.246.0.xx From: To:	DE (.54)	CH (.82)	IT (.129)	IT (.130)	LU (.145)	UK (.225)
AT (.20)	0.07	0.5	0.05	0.1	0.45	
AT (.22)		1	0.9	0.95	0.54	0.75
DE (.54)	1	0.9	0.8	0.7	1	0.6
DE (.55)	0.22	0.9	0.85	0.8	1	0.6
CH (.81)			0.75	0.65	1	0.7
CH (.82)	0.92		0.8	0.85	1	0.65
CH (.83)		0.9	0.8	0.75		0.6
ES (.100)	0.04	0.1			0.09	
ES (.101)	0.04	0.1				0.15
ES (.102)	0.14			0.05		0.05
IT (.129)	0.88	1		1	1	0.25
IT (.130)	0.88	1	1		1	0.65
IT (.132)		1	0.8	0.8	1	
IT (.133)				0.8		0.7
IT (.134)	1	1	1	1	1	0.6
LU (.144)					1	
LU (.145)	1	1	0.8	0.95		0.15
NO (.177)		0.3	0.65	0.8	0.45	0.8
NO (.178)	0.07		0.35	0.3		
NO (.184)	0.7	0.5	0.6	0.65	0.72	0.45
UK (.225)	0.77	0.5	0.6	0.65	0.72	
UK (.226)	0.33	0.44	0.5	0.5	0.81	0.95

Table 2.3: Probability that a SVC can be established

Discussion of results

- SVC set-up times are always well above the theoretical lower bound of one RTT and not neglectible.
- A huge number of connections failed permanently during the test period, but were known to work earlier. Since connections between the same sites using the same intermediate switches worked at the same time, it can be assumed that the end systems play an important role in those failures.

5.2.7 Major observations during the tests

Discrepancy between LAN and WAN

SVCs are being used successfully in local area networks already. Extension to the WAN is not easily possible as the LAN and WAN environments differ in some key aspects. Bandwidth in the LAN is relatively abundant and cheap, while it is scarce and expensive in the WAN. The ATM services used today in the LAN and WAN differ for that reason, even though the same physical infrastructure is used to carry both: While UBR SVCs are suitable for a WAN environment, they do not work over policed WAN VPs, where less than line speed is available for cost reasons. Furthermore, signalling is not currently supported by the JAMES infrastructure available for our tests. On the other hand, the CBR or VBR services used over WAN links play a less important role in the LAN due to the configuration overhead involved. We are left for the time being with two somewhat incompatible worlds. In our experiment, we try to expand the typical LAN service SVC to the WAN.

Relevance of Reshaping

Most, but not all of the switches used in our SVC network are capable of doing some form of traffic reshaping. This is required to ensure that the traffic contract on policed links is not violated. With sufficiently large buffers this can be used to shape reasonably well behaved UBR traffic into CBR VPs. No cells will be lost any longer due to policing, but excessive cells might have to be removed from the output queue. EPD, when available, makes sure that no fragments of packets get transmitted. This should work quite well with TCP/IP traffic, as the source will dynamically react to packet loss by adjusting the bandwidth usage. Excessive packet loss is thus prevented. In phase 3 testing we therefore eliminated or upgraded the switches without support for shaping from our SVC network and enabled shaping on all VPs towards policed VPs. Observations:

- Traffic policing is performed on a per VP basis as should shaping. But some of our equipment is only capable of shaping on physical interfaces. (With two spare interfaces per VP and a cable to interconnect them, per VP shaping can be emulated on those switches).
- Reshaping could cure neither the high failure rate to establish SVCs across the network nor the too high set-up times. Once established, SVCs proved to be suitable to carry general purpose TCP/IP traffic with good bandwidth utilisation and without cell loss due to policing.

Problems with tunneling of signalling messages

When establishing SVCs, the "network" side of a signalling tunnel communicates the VPI and VCI of the SVC to the "user" side of the tunnel. While the VPI on a VP is the same on both ends, they will generally differ on VP connection across multiple switches, and the user side has to replace the VPI number it receives in the signalling message with the VPI number it was received on. While Fore switches can handle this correctly, Cisco Lightstream switches can only operate signalling tunnels over VP connections, where both ends use the same VPI number. The JAMES contacts were helpful to reassign matching VPI numbers on both ends of JAMES VPs to overcome this implementation limitation.

VCI range mismatch

The "network" side of a signalling tunnel decides about the VCI to be used for a given SVC and communicates this decision together with other information to the "user" side of the tunnel. The VCI range is configurable on most ATM equipment. The range chosen by the "network" switch must be acceptable by the "user" switch or host. We observed that cisco switches do not always choose a VCI out of the range it is configured to choose from. The manufacturer has been informed about this.

Multiple SVCs between hosts

Normally only one SVC is set up between two hosts, when packets have to be exchanged between them. Between certain hosts, however, two SVCs are established. This is believed to violate the standards, but since no negative effects were detected, it has not been further investigated.

Available traffic classes

The only application at hand to make use of our SVC infrastructure is currently the ATM/AAL5/IP stack. Unfortunately, we could only configure UBR best effort SVCs on most equipment. Exceptions include cisco routers, where CBR/VBR SVCs can be used too. But none of our switches was able to handle SVC set-up messages with anything except UBR requests.

Study of SVC set-up times

Introduction and description.

Various measurements of SVC set-up times have been performed at INFN Milan to experimentally test the mechanism of set-up and release of SVCs using UNI 3.0 on a local and wide area network infrastructure.

The total time to establish a connection using SVC was measured in different configurations and an estimate of set-up times for various hardware is obtained. As James does not provide a native SVC service yet, we had to tunnel the signalling packets through CBR VPs using private switches, Cisco Lighstream 1010 or Fore ASX200. The complete network infrastructure has been described in a previous paragraph.

The time measurements were obtained using the standard ICMP utility 'ping' with a 64 bytes packet, which means that the delay time to transmit the packet is 0.4 milli-seconds in the case of 34 Mb/s and 5 micro-seconds in the case of 155 Mb/s..

Due to the fact that the UBR class of service is assigned to the SVC connections, the packet is sent at maximum link speed, but it was small enough to fit into the tolerance of the 2 Mb/s CBR VP.

When using a SVC based network, the first time an IP packet is sent, the virtual channel has to be created. After the connection is established at the ATM level, the first IP packet is sent. The following packets use the already existing channel, without suffering the SVC set-up time overhead. The time to set-up a connection can be thus defined as the delay the first packet experiences in addition to the actual IP RTT . It can therefore be computed by subtracting the IP RTT (i.e. the RTT of every packet except the first one) from the RTT of the first packet.

Repeating this process a large number of times in each case, it is possible to obtain a statistically significant estimate.

Between each measurement the connection was forced to be released, waiting a fixed amount of seconds to allow the release process to complete. Due to the simple test tools, the intrinsic precision of each measure was one millisecond, which had to be added to the statistical error.

All the measurements have been performed with the sending stations in single user mode or completely unloaded (i.e. by night).

Local measurements.

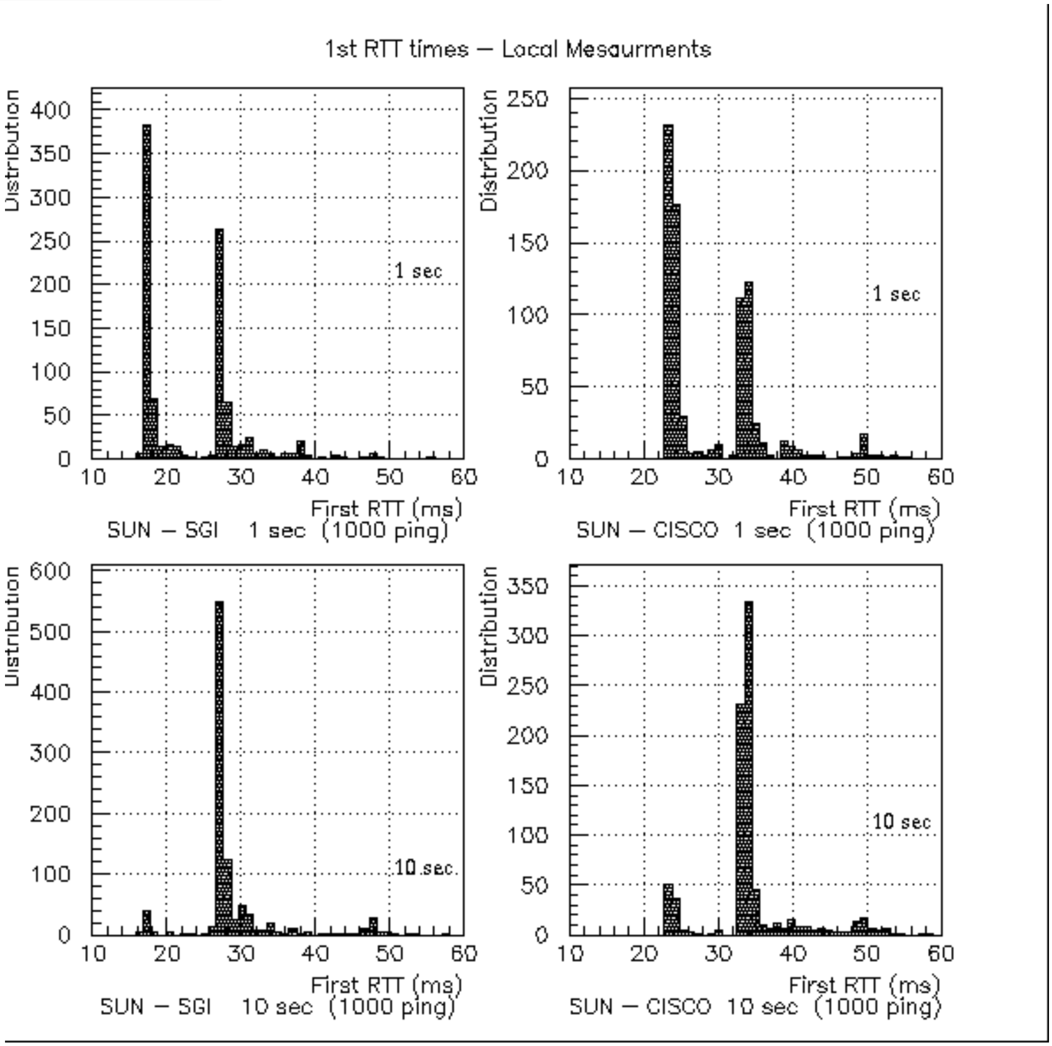


Fig. 2.1 Local frequency distributions, 1 and 10 seconds of sleep time.

Figure 2.1 shows a graph of the frequency distribution of the RTT for the first packet between the Sun and the SGI workstations and between the Sun and the Cisco 7507 using UNI 3.0, with 1 and 10 seconds of sleep time between each measure. Please remember that this RTT is the sum of the delay needed to establish a connection and the IP RTT of the first packet. Since from the tests done and considering the intrinsic error of one millisecond, the IP RTT is almost constant, we can consider the distributions reported in figure 2.1 as the distributions of the set-up time. As can be seen from figure 2.1, the distributions show clearly the presence of two peaks. Besides, with 10 seconds of sleep time, the second peak becomes predominant while the first one almost disappears. Hence the statistical population of the peaks depends on the sleep time between every measure. On the contrary the delta time between the peaks does not depend on the sleep time. Moreover, according to our measures, the delta time does not seem to depend on the receiving workstation either. Since the situation with 10 seconds of sleep time can be considered as the situation closer to the case in which every SVC connection is established independently from the presence or not of a previous virtual channel, we decided to calculate the set-up time, as described above, by subtracting the IP RTT from the RTT of the first packet averaged on the second peak. Please remember that the IP RTT has to be considered constant. Therefore:

```
Set-up time= RTT_first_packet (2nd peak) - IP RTT
```

The results we obtained calculating the set-up times in this way are reported in table 2.4.

Average times	Sun-Cisco	Sun-Sgi	Sun-Fore	Sgi-Cisco	Sgi-Sun	Sgi-Fore
RTT first packet (1st peak)	23 ms	17ms	20 ms	21 ms	18 ms	18 ms

RTT first packet (2nd peak)	34 ms	27 ms	31 ms	34 ms	29 ms	29 ms
Delta between peaks	11ms	10 ms	11 ms	13 ms	11 ms	11 ms
IP RTT	1 ms	1 ms	1 ms	1 ms	1 ms	1 ms
Set-up time	33 ms	26 ms	30 ms	33 ms	28 ms	28 ms

Table 2.4 Local measurements.

Wide area network measurements

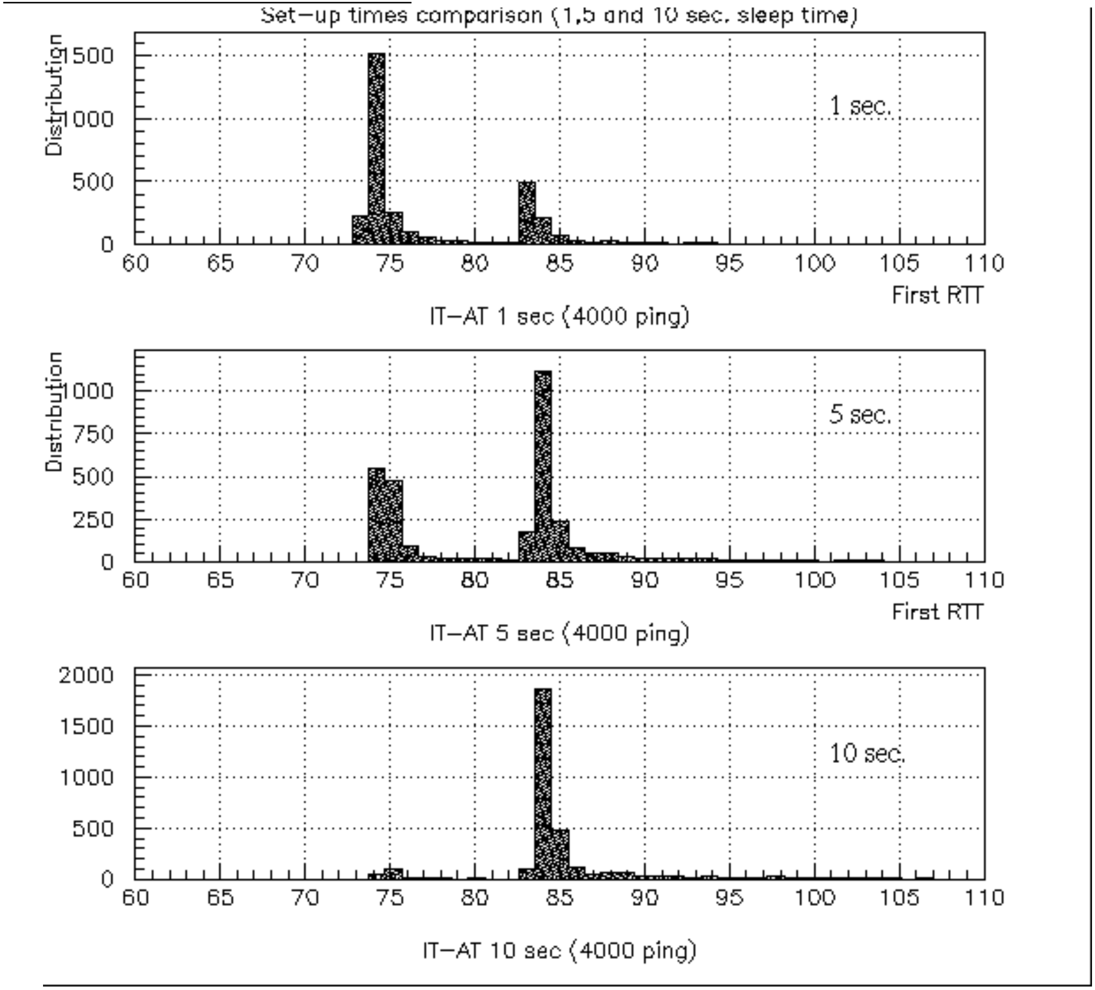


Fig. 2.2 Frequency distrib. between IT (193.246.0.129) and AT (193.246.0.22) at 1, 5 and 10 sec. of sleep time.

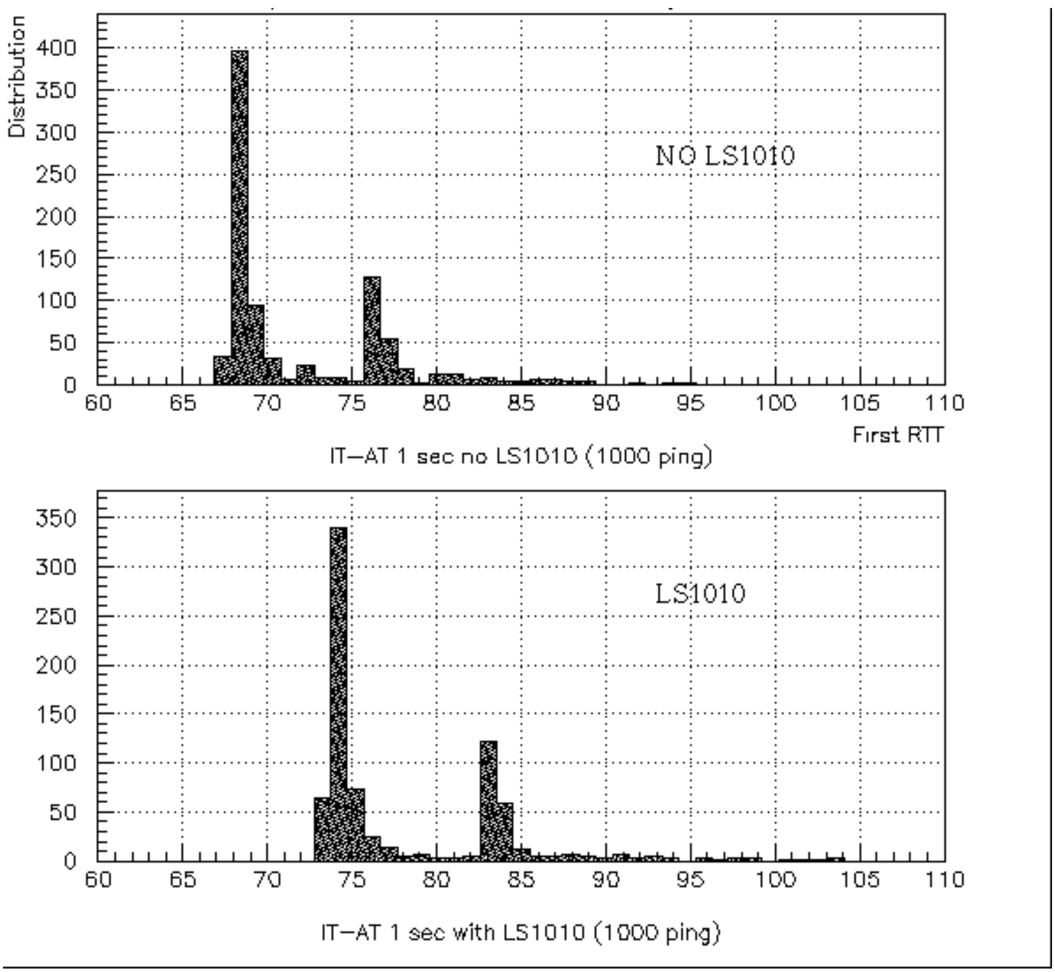


Fig. 2.3 Frequency distributions between IT (193.246.0.129) and AT (193.246.0.22) with and without LS1010 on the Italian side.

In figure 2.2 the frequency distributions of the set-up time between the Sun in Italy (193.246.0129) and the Cisco 7010 in Austria (193.246.0.22) at 1, 5 and 10 seconds of sleep time are reported. Like the case of the local area network measures, the migration of the statistical population from the first to the second peak as the sleep time increases and the independence of the delta time between the peaks from the sleep time can be clearly seen. Besides, from the tests done (see tables 2.5 and 2.6) it seems that the delta time is not a function of the distance between the sending and the receiving workstation and on the type of end-workstations either. Therefore the delta time, and the presence of two peaks, must be due to some mechanism intrinsic to the ATM network.

In figure 2.3 we report the frequency distribution of set-up times from our Sun to the Cisco 7010 in Austria in the two cases where the Cisco Lighstream 1010 was or wasn't present on the Italian side. The sleep time is one second. These two distributions are equal except for a shift, in the second case, of 6 milliseconds. This shift could be therefore assumed as the delay time introduced by a Cisco Lightstream 1010 in the set-up of a SVC.

Calculating the set-up time as described in the previous paragraph, in tables 2.5 and 2.6 we report our measurements in a wide area network. In table 2.5 the measures performed from our Sun to different destinations are reported, while in table 2.6 the measures from our Silicon to almost the same destinations are reported. The * indicates the presence of the Cisco Lightstream 1010 on the Italian side, otherwise only the Fore ASX200 was present on this side of the tunnel.
Average times

	CIS7010 (AT)*	SUN (CH)*	SUN (DE)	CIS7010 (DE)	CISCO (BE)
RTT first packet (1st peak)	74 ms	42 ms	121 ms	125 ms	207 ms

RTT first packet (2nd peak)	84 ms	51 ms	129 ms	133 ms	216 ms
Delta between peaks	10 ms	9 ms	8 ms	8ms	9 ms
IP RTT	18 ms	7 ms	47 ms	47 ms	74 ms
Set-up time	66 ms	44 ms	82 ms	86 ms	142 m

Table 2.5. Measures from SUN(IT) (193.246.0.129). * means LS1010 was present.

Average times	CISCO7010 (AT)	SUN (CH)	SUN (DE)	CISCO7010 (DE)	CISCO (BE)
RTTfirst packet (1st peak)	67 ms	35 ms	119 ms	124 ms	206 ms
RTT first packet (2nd peak)	75 ms	43 ms	127 ms	132 ms	215 ms
Delta between peaks	8 ms	8 ms	8 ms	8 ms	9 ms
IP RTT	18ms	7ms	47 ms	47ms	74 ms
Set-up time	57ms	36 ms	80 ms	83 ms	141 ms

Table 2.6 Measures from SGI(IT) (193.246.0.130).

The percentage of failures reported during our tests was variable between 13 and 15 percent. Since these failure rates were present in the local tests made with the Cisco Lightstream 1010 but not in those made only with the Fore ASX200, we charge these failures to some negotiation mismatch between Cisco and Fore software (for example in the VCI bits range).

Estimate of the set-up times of every single machine involved.

From our local measurements, as we know the exact configuration of the network involved in each test, it is be possible, as a simple exercise, to determine the SVC set-up time of each single network element (end-stations and private switches). In fact considering the set-up time as the sum of the set-up times of the end-workstations plus the set-up time due to the switch it follows that:

```
Set-up Time=Set-up(Host<->Switch)+Set-up(Switch)+Set-up(Switch<->Host)
```

Making the assumption that the set-up times in initiating or in receiving the connection is identical and that the switch counts two times (3 if it is the destination), we could solve a linear system of 3 equations in 3 unknown quantities (the single network element SVC set-up times) based on the measures reported in table 2.4. Besides, using also the WAN estimate of the delay in the set-up time introduced by a Cisco LightStream1010 (6 ms), we could calculate the delay time introduced by the Cisco 7507. Table 2.7 summarises the results obtained which have to be taken with an error of 1ms.

LOCAL HOSTS	Sun	Sgi	Fore*	Cisco 7507	LS1010*
Set-up time (ms)	6 ms	4 ms	8 ms	6 ms	3 ms

Table 2.7 Single hardware set-up times. * means the set-up time is per switch interface.

Conclusions

From our measurements a few simple deductions can be drawn:

- The IP round trip times are longer then expected, i.e. the time needed by packet to travel across the SVC network is much longer then the time required by an electric signal to travel along the same distance. This means the actual structure of the network plays a fundamental role.
- The set-up times themselves are not negligible. In particular the set-up times in a wide area network, being the sum of the initiator delay, plus the receiver delay, plus the switching delay multiplied by the number of switches, plus a certain amount of time due to the exchange of ATM signalling packets may sum up to a fraction of a second on a

complex network. The estimate of the exact SVC set-up times is difficult due to the overlap in time of the exchange of the signalling packets between the network nodes.
- The statistical distributions of the set-up times is a function of different parameters, mainly the sleep time and the load of the machines (we made measurements both in single and in multi user mode and obtained different distributions, which the higher was the load of the sending node, the greater was their difference).

Some irreproducible results

Some of our findings could not be properly reproduced, nor could we find conclusive explanations for them. But we would like to state our observations nevertheless:

- Sometimes during our ping-tests, ICMP redirect messages were received, indicating that packets were sent to the wrong host. Most probably, the sending host chose the wrong SVC to reach the destination (observed on the UK workstation).
- It could be observed that an ATM end system and its switch were not in sync with respect to the number of SVCs that should be in place between them. This could have been caused by either the switch or the ATM end system discarding signalling messages (observed twice by accident between the UK host and switch). The SVCs the switch was no longer aware of did no longer work and packets to those destinations sent by the host were dropped on the switch.
- The host suddenly refuses to send set-up messages to certain hosts. This could be the result of certain reject messages received from the network, but could not be investigated further due to the lack of ATM analyser equipment (observed several times on the UK host).
- Spontaneous reboots of ATM equipment could be observed in different sites involved in SVC testing, but the conditions leading to this behaviour require ATM analyser facilities for further investigation.

5.2.8 Relevance for service and outlined migration to service

The TEN-34 backbone will initially consist of CBR PVPs terminating its single VC on IP routers at either end. Resilience protecting against link failure is reached by re-routing on the IP layer.

In a more advanced set-up, the connections between the same or similar set of routers could be done by ABR SVCs resulting in a partial or full mesh between those routers. Depending of available services from the WAN link provider, either tunnelling or native SVC will be used. This set-up provides some major advantages with respect to the initial one:

- Switches introduce less transmission delay than routers. Traffic between TEN-34 sites will transverse fewer routers, thus resulting in reduced transmission delay.
- The same network infrastructure can be used to home additional services, i.e. native ATM services.
- Re-routing on the ATM layer is transparent to the IP layer and will therefore not produce any route flaps in IP routing, as it is the case with PVCs.
- Further experience must be gained in the following fields prior to deployment:
    - dynamic ATM routing
    - Management of ATM switched networks
    - SVC with ABR
    - Native SVC

Our tests show, that a signalling infrastructure based on the equipment at hand right now, is not yet stable enough to support such an infrastructure in a production environment.

5.2.9 Test-related problems and general comments

- The JAMES procedures and the overhead involved in setting up new VPs between our sites proved to be a too complicated and lengthy process to be able to order VPs at short notice as needed. Therefore, wherever possible, the "overlay network" was used for our tests.
- A general comment on the complexity of ATM: Despite we considered ATM to be a complex technology, we almost always underestimated the effort required to setting up networks based on ATM.

5.2.10 Further studies

- Trying to understand the high increase of set-up times over WAN links.
- The decision to use SVC tunnelling instead of native SVC was mainly due to the fact that JAMES did neither offer support for signalling, nor decide about an addressing scheme yet. As soon as those two issues are resolved, native SVC should be tested directly instead of tunnelling.
- The only application tested so far was the ATM/AAL5/IP stack. Other applications should be considered as well.
- Our switches support currently only switching of UBR VCs. Other traffic classes should be considered too as they become available.
- Static IISP routing (aka PNNI phase 0) was used throughout our tests, but PNNI should be tried too.
- Reliability and delay problems were detected when establishing SVCs. The reasons for this behaviour are currently not properly understood and require more investigation.

5.2.11 Annex

Static IP to NSAP mapping and NSAP prefix table

```
# Mapping between IP and NSAP addresses for SVC testing over JAMES
# and NSAP prefixes used on involved switches
#==================================================================
#  last update: 25/03/97 CG
#
# IP address            NSAP address
#
# ACONET (AT)
prefix:                 39.040F.5404.0101.XXXX.XXXX.XXXX.XXXX.XXXX.XXXX.XX
193.246.0.20            39.040F.5404.0101.0001.9999.0002.0020.EA00.0B22.00
193.246.0.22            39.040F.5404.0101.0001.9999.0001.9999.9999.9901.50
#
# ULB/STC (BE)
prefix:                 39.056F.XXXX.XXXX.XXXX.XXXX.XXXX.XXXX.XXXX.XXXX.XX
193.246.0.32            39.056F.0000.0000.0000.0000.0000.0001.9324.6032.01
#
# Belgacom (BE)
prefix:                 47.0005.80FF.E100.0000.F215.100F.XXXX.XXXX.XXXX.XX
193.246.0.40            47.0005.80FF.E100.0000.F215.100F.0020.4815.100F.00
#
# DFN/RUS (DE)
prefix:                 39.276F.3100.0110.0000.0001.XXXX.XXXX.XXXX.XXXX.XX
193.246.0.54            39.276F.3100.0110.0000.0001.0003.0020.4806.0989.01
193.246.0.55            39.276F.3100.0110.0000.0001.0003.1111.1111.1102.04
#
# RCCN (PT)
prefix:                 39.620f.0000.0000.0000.0000.0000.XXXX.XXXX.XXXX.XX
193.246.0.73            39.620f.0000.0000.0000.0000.0000.0020.4806.84b9.01
193.246.0.74            39.620f.0000.0000.0000.0000.0000.0020.481a.3714.01
#
# SWITCH (CH)
prefix:                 39.756F.1111.1111.7001.0001.1002.XXXX.XXXX.XXXX.XX
193.246.0.81            39.756F.1111.1111.7001.0001.1002.1932.4600.0081.01
193.246.0.82            39.756F.1111.1111.7001.0001.1002.1932.4600.0082.01
193.246.0.83            39.756F.1111.1111.7001.0001.1002.1932.4600.0083.01
#
# REDIRIS (ES)
prefix:                 39.724F.10.010001.0001.0001.0001.XXXX.XXXX.XXXX.XX
193.246.0.100           39.724F.10.010001.0001.0001.0001.0020.481A.1E5E.01
193.246.0.101           39.724F.10.010001.0001.0001.0001.1932.4600.0101.00
193.246.0.102           39.724F.10.010001.0001.0001.0001.0020.4806.225B.00
#
# INFN (IT)
prefix:                 39.380F.XXXX.XXXX.XXXX.XXXX.XXXX.XXXX.XXXX.XXXX.XX
193.246.0.129           39.380F.0000.0000.0000.0000.0000.0019.3246.0129.01
193.246.0.130           39.380F.0000.0000.0000.0000.0000.0019.3246.0130.01
193.246.0.131           39.380F.0000.0000.0000.0000.0000.0019.3246.0131.01
193.246.0.132           39.380F.0000.0000.0000.0000.0000.0019.3246.0132.01
# 193.246.0.133         39.380F.0000.0000.0000.0000.0000.0019.3246.0133.01
```

```
193.246.0.134           39.380F.0000.0000.0000.0000.0000.0020.4815.15A9.01
# 193.246.0.135         39.380F.0000.0000.0000.0000.0000.0019.3246.0135.01
#
# RESTENA (LU)
prefix:                 39.442F.XXXX.XXXX.XXXX.XXXX.XXXX.XXXX.XXXX.XXXX.XX
193.246.0.144           39.442F.0000.0000.0000.0000.0001.0020.481A.1D5B.00
193.246.0.145           39.442F.0000.0000.0000.0000.0001.0020.4806.221E.00
#
# UNINETTT (NO)
prefix:                 47.0023.0100.0005.XXXX.XXXX.XXXX.XXXX.XXXX.XXXX.XX
193.246.0.177           47.0023.0100.0005.2000.0001.0101.1034.1034.1034.01
193.246.0.178           47.0023.0100.0005.2000.0101.0120.0800.093d.0641.00
193.246.0.184           47.0023.0100.0005.4000.0001.0101.0800.093d.063c.01
#
# UKERNA (UK)
prefix:                 39.826F.1107.2500.10XX.XXXX.XXXX.XXXX.XXXX.XXXX.XX
193.246.0.225           39.826f.1107.2500.1000.0000.0000.0020.4806.1ff1.00
193.246.0.226           39.826f.1107.2500.1000.0000.0000.0020.481a.2e52.01
```

5.2.12 References

[1] ATM Forum, "ATM User-Network Interface Specification Version 3.0", 1993

[2] ATM Forum, "ATM User-Network Interface Specification Version 3.1", 1994

[3] J. Heinanen, "Multiprotocol Encapsulation over ATM Adaptation Layer 5", RFC 1483, Telecom Finland, July 1993

[4] M. Laubach, "Classical IP and ARP over ATM", RFC 1577, Hewlett-Packard Laboratories, January 1994

[5] M. Perez et al., "ATM Signaling Support for IP over ATM", RFC 1755, USC/Information Sciences Institute, February 1995

5.3 Classical IP and ARP over ATM

5.3.1 Experiment Leaders

Simon Leinen, SWITCH, CH
Ramin Najmabadi Kia, ULB/STC, BE

5.3.2 Introduction

In Classical IP over ATM as defined in [RFC1577], a specialised variant of ARP server is used to resolve layer-three (IP) addresses to layer-two (ATM NSAP or E.164) addresses. The main difference to traditional ARP is that - because ATM lacks a broadcast facility - there is a single designated ATMARP server whose layer-two address has to be configured statically in each client.

5.3.3 Protocol Operation

Classical IP (CLIP) is based on ATM Switched Virtual Circuits (SVCs). It is only defined within a Logical IP Subnet (LIS).

When a CLIP node wants to send an IP packet to another CLIP node on the same LIS, and no SVC been the two nodes has been established yet, the sending node has to request an ATM SVC to the receiver. For this purpose, it needs to know the receiver's ATM address. Unless the mapping is already in the cache, it queries the ATMARP server.

Likewise, when a CLIP node receives an SVC connection request from another node, it uses an Inverse ARP (InARP) request to the ATMARP server to find the protocol address of the sender.

Communication between CLIP nodes and the ATMARP server is performed using AAL5/SNAP over a regular SVC, and the ATM address of the ATMARP server has to be configured statically in each node on the LIS. This SVC can also be used to carry IP traffic between a node and the node running the ARP server.

5.3.4 Experiment Setup

Building on the configuration for the SVC tunneling experiments, another range of network addresses (193.203.225.0/24) was reserved for this experiment. Volunteers had to configure an additional ATM sub-interface in a CLIP/ATMARP configuration on their nodes. An ATMARP server was configured on a Cisco router at the University of Linz in Austria, which was used by all participants. The only addresses that had to be configured on each participating interface were:

- the local IP address
- the ESI of the local NSAP address
- the NSAP address of the ATMARP server

This compares quite favourably with the setup for the SVC tunneling experiment, where every participant needed a complete table of IP/NSAP mappings for all other interfaces.

5.3.5 Observations

Using an ATMARP server didn't introduce any new instabilities for the participants. However, problems with the SVC tunneling network could prevent potential participants from contacting the ATMARP server, which would make all communication within the LIS impossible, even though some destinations would be reachable on the ATM level. On the other hand, static IP-to-NSAP address mappings aren't necessary when ATMARP is used, removing another common source of errors and maintenance effort.

5.3.6 Timing Results

The following table compares response times for ICMP echo requests ("pings") within the same LIS, on the one hand using an ATMARP server, on the other hand using statically configured IP-NSAP address mappings. The first packet takes a bit longer to respond to using ATMARP, because the ATM server has to be contacted by the sender (ARP request) and/or responder (InARP request). For subsequent packets, the response time was the same in both setups, except for differences due to the SVC tunnel topology that had changed between both tests. Ideally all the experiments would have been done at the same time. For technical and organisational reasons this was not possible, so that it is problematic to compare the setup times.

The SVC between an ATMARP client and the server is usually kept active permanently, so calls to the ATMARP server are not included in the timings below.

Notes: All timings were taken from a Sun in Switzerland ("CH2"). All times are expressed in units of milliseconds (ms).

dest	with ATMARP		without ATMARP		notes
	1st	nth	1st	nth	
CH1	NR	1	NR	1	First packet is always dropped
CH2	0	0	0	0	
CH3	NR	1	11	1	
AT1	13	13	55	13	ATMARP server, so no connection setup overhead with ATMARP
NO1	386	84	242	86	
NO2	619	78	299	90	
LU1	200	67	223	68	
UK1	914	123	573	142	
UK2	846	124	406	132	

The table shows some inconsistencies of the response times, which are due to different signaling load on the switches and routers over the experiment, differences in topology of the network, and differences in implementations. It was not possible with the equipment and time available to resolve these inconsistencies.

Another experiment consisted of setting up two CLIP LISes, using the same machine as the ARP server. Between the LISes, normal layer-three routing was used. This worked as expected, with no ATMARP information leaking between the two subnets.

5.3.7 Conclusion

ATMARP works quite well as an address resolution protocol mapping IP to ATM NSAP addresses. Its use yields an extremely simple configuration for an IP subnetwork over an ATM SVC infrastructure. Protocol overhead is very small and only noticeable on new connections.

The dependency on a single ATMARP server is a severe drawback, in particular in a WAN setting. On a LAN, this may be acceptable because the server can be run on a system whose functioning is vital to the network anyway (such as a central server or switch). But on a WAN, ATM-level connectivity problems cause the ATMARP server to be reachable for some parties, making all communication impossible, including to other parts that can still be reached over the ATM network.

The NBMA Next Hop Resolution Protocol [NHRP] alleviates the problem by allowing smaller LISes and permitting layer-2 connectivity outside the LIS. Having multiple redundant address resolution servers necessitates a synchronization protocol such as Server Cache Synchronization Protocol [SCSP]. ATMARP and NHRP clients can coexist as described in [CLIPNHRPTR], but ATMARP clients will use layer-three routing to reach nodes outside the LIS.

This problem has been recognised by the IETF, and in the framework of the evolution of the Classical IP model, there will be a proposal on how SCSP can be used to keep multiple ATMARP servers consistent. The advantage of such a solution would be that existing ATMARP clients would function with no modification, whereas NHRP has not been widely implemented in client stacks yet.

5.3.8 References

[RFC1577] M. Laubach, Hewlett-Packard Laboratories, "Classical IP and ARP over ATM", January 1994

[SCSP] James V. Luciani, Grenville Armitage, Joel Halpern, "Server Cache Synchronization Protocol (SCSP)", Internet-Draft, November 1996 (word in progress)

[NHRP] James V. Luciani, Dave Katz, David Piscitello, Bruce Cole, "NBMA Next Hop Resolution Protocol (NHRP)", Internet-Draft, March 1997 (work in progress)

[CLIPNHRPTR] James V. Luciani, "Classical IP to NHRP Transition", Internet-Draft, October 1996 (work in progress)

5.4 IP routing over ATM with NHRP

5.4.1 Participants

Olav Kvittem(leader), Vegard Engen - Uninett, Simon Leinen - Switch, Guenther Schmittner - University of Linz, Robert Stoy - University of Stuttgart and Celestino Tomas - RedIris.

5.4.2 Summary of results

This projects has set up an experimental IP over ATM network using the Next Hop Resolution Protocol(NHRP). The network spanned 5 countries and used ubiquitous ATM SVC-tunneling so that any pair of participants could make ATM-connections to each other. NHRP was demonstrated to work and gave all participants direct IP/ATM connections without the need for manual mapping tables or a centrally managed IP subnet.

5.4.3 Dates and phases

The project was prolonged for 4 months in order to get broader experiences in operational requirements. There is a delay in about one month in starting the experiment phase partly due to dependence on SVCs.

Revised plan	dates	results
1.Investigation	96-07 - 96-10	
2.Initial experiments	96-10 - 96-12	detailed pilot documentation
2.Pilot experiment	97-01 - 97-03	operational infrastructure
3.Reporting	97-03 - 97-04	report

5.4.4 Network infrastructure

The project used the ATM SVC infrastructure set up by the SVC-project.

5.4.5 Results and findings

Background

An IP-system at the edge of an ATM-network needs to find for a destination IP-address the ATM-address for the optimal next hop over the ATM-network so that it can set up a call there. A partial solution to this problem is the the ATM ARP in RFC1577 (Classical IP over ATM) which solves the problem for one IP subnet. This does not scale to large multiorganisation networks. The Next Hop Resolution Protocol (NHRP) proposes a solution for shortcutting subnetbased routing so that one can minimize the number of hops through the same ATM cloud.

Given an european academic ATM-based backbone with possibly more than 40-50 nodes, NHRP might be the way a pan-euroepean academic IP-network could be practical to set up. With statically set up connections the network would be tedious to maintain and lead to incomplete direct connectivity and thus inefficient use of network resources. With NHRP one could hope for automatic setup of connections to new nodes with a traffic interest. The same problem exist perhaps to an even bigger scale in national academic networks.

Clients and servers

The current status of the development of NHRP at IETF is that the protocol is under consideration by the IESG as a proposed standard. This means that the protocol is fairly stable.

There is however still few implementations available. There has been an implementation for Cisco routers available for a while. This was chosen for the tests. There is also one for a workstation, but for an older incompatible version of the protocol.

SVC infrastructure

The NHRP operation is dependent on having a ubiquitous SVC-connectivity among the participants forming a logical NHRP cloud over the TF-TEN ATM VP overlay network. Such an infrastructure has been prepared by the ATM SVC-project. However the NHRP project copied that setup putting in its own VC's in order not to interfer with the other experiments like SVC and ATM-ARP.

NHRP operation

The routers at the edge of the ATM-cloud will act as NHRP servers. There need to be a initial connectivity between routers/hosts on the IP level so that NHRP can work. This can be a slow indirect path. The initial predefined VC-connections defined are that each country connects to one of two interconnected centers in Germany and Austria.

The NHRP servers will have the same network-id, that will tell them that they are on the same ATM-cloud when receiving a NHRP-request, and may return info about their ATM-address to the requestor. A NHRP request will be sent when a predefined amount of packets has been sent towards a destination. The request will be passed along to NHS servers
on the ATM-cloud until no further downstream NHS-servers are available. The egress router from the ATM-cloud (the router at the exit of the ATM network) will then return his ATM-address to the ingress router.

Tests

A demonstration of the basic behaviour of the Cisco implementation is shown in the following simple test:

A and B have an IP-link with the ATM address of each other as well as B and C, but A and C do not know how to contact each other.

- A sends echo packets towards C via the default route to B.
- A brings a SVC to B to serve that traffic.
- B sends the packet on to C and brings up an SVC to do that.
- C returns the packets to A via B
- A sends a NHRP request to C via B after some packets. The NHRP packet contains A's ATM-address and
- C tries to setup a direct connection via the ATM-address from the request but fails due to SVC-problems
- C responds to A via B with it's ATM-address
- B receives the reply and sets up a SVC to A

This experiment was performed beween Austria, Switzerland and Norway (ABC) and the roundtrip time with the A-B-C path was 108 milliseconds, while it was about 76 ms with the A-C VC.

Conclusions

This simple experiment has demonstrated that and how the NRHP basic functions works. The implementation is still largely untested and we experienced router crashes, routing tables flushes and looping SVC-control processes during testing. There were also some problems on the ATM SVC-level that are mentioned in the SVC-experiment

There is also some functionality missing in the ATM implementation of the router, like queing up packets while waiting for a call to be set up. As it is now, packets coming in are lost until a call is active.

This version of NHRP only supports lookup of addresses directly reachable from the egress router. This means that to make transit traffic beween the networks behind the respective connections flow on the NHRP connection one must use normal routing on top. NHRP would be more useful in a backbone with such an extension (NHRP-R2R).

The present ATM network (JAMES + NRNs) does not support any means of resource control in the network besides static allocations. Due to inherent properties of ATM the packet loss can be disastereous when a link is saturated using Unspecified BitRate(UBR). Setting up a large number of unrelated NHRP UBR VC's in a not controlled resource environment is not recommeded. NHRP does not have any resource reservation mechanisms, so one would have look to ATM mechanisms like Packet Discard, Available BitRate Services and resource reservation, or to higher level like RSVP.

5.4.6 Relevance for service and migration suggestions

The present status of standardisation and implementations is unstable and not yet mature for production environments. However the NHRP mechanism as such could be become a practical way of engineering an IP overlay network on a potential large scale integrated European Academic ATM-backbone.

5.4.7 Test related problems and general comments

There were some initial problems getting ATM SVCs to work. The quality of the implementations used destabilized the participating systems, so use of production routers for experiments shold be done with care.

5.4.8 Further studies

It is highly recommended that this project continues with the targets of advancing on the above mentioned issues like a larger scale pilot, transit routing interaction and ATM resource control mechanisms.

5.4.9 References

[1] SVC and ARP test for TF-TEN

[2] Braden, R., Zhang, L.,: Resource ReSerVation Protocol (RSVP) -- Version 1 Functional Specification, Work in Progress, 1996

[3] Integrated Services Model, RCF 1633, 1995

[4] Laubach, M.,: Classical IP and ARP over ATM, RFC 1577, January 1994

[5] Schill, A., K¸hn, S., Breiter, F.: Internetworking over ATM: Experiences with IP/IPng and RSVP, 7th Joint European Networking Conference, Budapest, Hungary

5.5 European ATM Addressing

5.5.1 Experiment Leader

Kevin Meynell, UKERNA

5.5.2 Summary of Results

Most NRNs have decided they wish to use NSAP addresses for ATM signalling. All the PNOs however, have decided to use E.164 addressing. Whilst NSAP address formats are well defined, there are still no standards for deriving E.164 addresses from these. Until these are published, the scope for extending signalling across the JAMES network is restricted.

5.5.3 Participants

UKERNA, University of Edinburgh, UNINett, DANTE, ACOnet, SURFnet

5.5.4 Results and Findings

The aim of this project was to devise an ATM addressing scheme for European NRNs that would allow experiments with UNI signalling and routing services. It was also hoped that a universal scheme would allow the scope of the JAMES experiments to be easily expanded, and avoid a lot of re-configuration work in the future.

Most NRNs have indicated they would prefer to use NSAP addressing as this provides the fine address resolution they are likely to require. As various NSAP formats are well defined, it is really only necessary for each NRN to obtain an NSAP prefix from the ISO National Member Authority for their country (in the UK this is the British Standards Institute). The NRN may then allocate the undefined octets in a manner that suits it is topology/organisational structure. JANET, the UK NRN has devised a scheme that could possibly be adapted by other NRNs (http://www.ed.ac.uk/~george/ukac-index.html).

Most of the European PNOs however, have indicated they will be using E.164 addressing, the ITU standard relating to international ISDN numbering. Consequently, this means there must be a method for NSAP addresses to traverse the PNO-provided network.

ATM Forum standards state that where a call originates from, and is destined for, networks supporting NSAP addresses, the NSAP address may be carried in the E.164 sub-address field over an E.164 network. The E.164 address (Called Party Number) required for transit must be derived from the NSAP address at the gateway between the two networks. Where a call originates from a network supporting NSAP addresses and is destined for a network only supporting E.164, the Called Party Number will be coded as an NSAP-formatted E.164 address.

Unfortunately, there are not any standards for this and translation appears to have been left to the switch suppliers to implement. The only switch supplier known by the author to be working on a solution is Cisco and this is proprietary.

Another problem is the differences in field length between E.164 and NSAP addresses, and the fact that some telecommunications switch manufacturers do not support the full E.164 field length. This could conceivably mean that parts of an NSAP address would be discarded when entering an network only supporting E.164. Indeed, the PNOs themselves are not yet sure how to proceed on these issues.

The ATM Forum and ITU are currently working to define some standards in these areas, but nothing firm has been published. Until this happens, which is unlikely to be until next year, further progress will be inevitably restricted.

Nevertheless, it is not currently an issue for the SVC experiments over JAMES as they are being tunnelled over VPs. The NRNs should still also be able to determine their NSAP addressing schemes to use, which would allow real values to be assigned to their equipment (as JANET as done). Indeed, this would benefit their own internal ATM networks.

The following table provides a summary of the known address schemes that will be used by European NRNs and the JAMES partners:

Country	NRN	PNO
Austria	NSAP DCC	E.164
France	NSAP DCC*	E.164
Germany	NSAP DCC	E.164
Italy	NSAP DCC	No decision
Netherlands	NSAP DCC*	E.164
Norway	NSAP ICD	No decision
Spain	NSAP DCC	E.164
UK	NSAP DCC*	E.164 NSAP (interim)

* denotes an offical scheme has been published

5.5.5 Further studies

Cisco has recently introduced address translation for it's Lightstream ATM Switches and it should be possible to start testing this. It is also necessary to continue to monitor progress on the standards relating to addressing at the ATM Forum and the ITU.

5.5.6 Bibliography and references

[1] SVC Tests
[2] ARP Tests
[3] Howat, G; JANET ATM Addressing Scheme; University of Edinburgh; 1996
[4] Howat, G; ATM Addressing Discussion Paper; University of Edinburgh; 1996
[5] Olsen, K; The UNInett Addressing Scheme; University of Oslo; 1995
[6] Reijs, V; ATM Addressing; SURFnet; 1996
[7] ATM Forum Technical Committee; ATM Forum UNI 3.1 Specification; ATM Forum; 1994
[8] ATM Forum Technical Committee; ATM Forum UNI 4.0 Specification; ATM Forum; 1996
[9] ATM Forum Technical Committee; ATM Forum PNNI 1.0 Specification; ATM Forum; 1996

5.6 ATM Network Management

5.6.1 Experiment leader:

Zlatica Cekro, University of Brussels, ULB/STC

5.6.2 Summary of results

Experience is gained on the management access to the NRN ATM edge devices, on monitoring and statistics collection using ATM related MIBs (Management Information Bases). SNMP (Simple Network Management Protocol) versions 1 and 2 were used. OAM (Operations, Administration and Maintenance) flows F4 and F5 (defined in ITU-T I.610) for the ATM layer Loopback connectivity detection were activated and tested. A Management Platform based on SunNet Manager-SunNet Domain Manager version 2.3 on Solaris 2.4 from the University of Brussels was used. The management platform enabled "monitoring information" i.e. read only class of service for nine NRN ATM switches and three routers with ATM interfaces. It was available for all participants in the tests through the remote X-window sessions over the Internet.

The transport links between the management platform and NRN ATM devices were realized through the operational Internet service with one link over switched ATM connection. The test results concern the following:

SNMPv1 and SNMPv2 based agents are widely implemented at the tested NRN edge devices: CISCO LS1010, CISCO LS100, FORE ASX200, UB GeoSwitch, CISCO routers with ATM interfaces. ATM based standard MIBs like ATM MIB (IETF RFC 1965) and ATM FORUM UNI MIB are widely supported by tested ATM switches. Very rich proprietary ATM MIBs at FORE and CISCO were tested. OAM F4 and F5 Loopback flows (ITU-TS I.610) were tested at the CISCO ATM switches.

5.6.3 Participants

For the management functions analysis and evaluation of management services: ACOnet (AT), ULB/STC (BE), CERN (CH), SWITCH (CH), DFN (DE), NORDUnet (SE and NO), SURFnet (NL), RedIRIS (ES), GARR (IT), UKERNA (UK).

For the phases of intensive testing: NORDUnet (SE and NO), GARR (IT), ULB/STC (BE), SURFnet (NL), ACOnet (AU), SWITCH (CH), DFN (DE), UKERNA (UK).

5.6.4 Dates et phases

In general phases started as it was proposed but dates were delayed for one month. Network Management tests were performed continuously over the period of September '96 - April '97.

5.6.5 Network infrastructure

The existing ATM Overlay network (User Information transport network) with NRN ATM edge devices was used for the tests. As management transport network we used two infrastructures: The Internet and the ATM Overlay network. No special configuration of ATM Overlay network for the management tests was required.

NRN ATM equipment participated in the tests included:

- NORDUnet: Norway, Oslo, CISCO ATM switch LightStream100 (LS100) and CISCO router,
- GARR: Italy, Milan, FORE ATM switch ASX200 and CISCO ATM switch LightStream1010 (LS1010),
- ULB/STC: Belgium, Brussels, CISCO ATM switch LS100 and CISCO router 7010,
- ACOnet: Austria, Linz, CISCO ATM switch LS100,
- SWITCH: Switzerland, Zurich, CISCO ATM switch LS1010 and CISCO router,
- SURFnet: Nederlands, Twente, UB GeoSwitch 155,
- DFN: Germany, Stuttgart, CISCO ATM Switch LS1010,
- UKERNA: United Kingdom, London, FORE ATM switch ASX200.

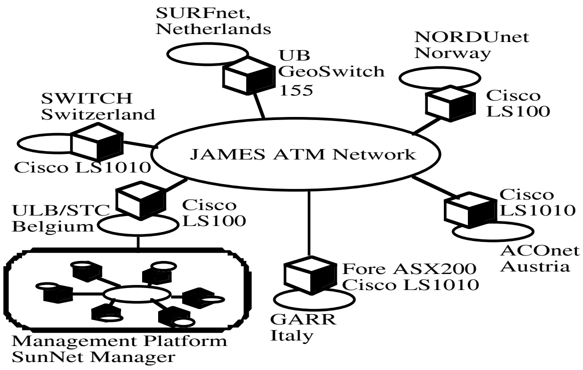


Figure 6.1: Infrastructure used for ATM Network Management tests

5.6.6 Local infrastructure

SunNet Manager-SunNet Domain Manager version 2.3 on Solaris 2.4 was used as a Management Platform for Network monitoring. It was connected both to the Overlay ATM Network and Internet network for the tests.

5.6.7 Hardware/software

On the NRN edge devices: The releases of software which supported the latest standards were required; like MIB II and SNMPv2 and ATM Forum MIBs.

On the Management platform: The SNMPv1 and SNMPv2 management request support, all standard MIBs and all proprietary MIB agents supported by the NRN edge devices were available.

5.6.8 Results and findings

Analysis of management possibilities in NRNs and PNOs

The work on the analysis has been realized in July and August 1996. The investigation of management possibilities at the NRNs and PNOs side resulted in modification of our initial test scenario from D 11.1, Version 2, July 1996. Further on, these two scenarios are described in more details.

Initial test scenario

The principal tests ware based on the ATM Forum Specification M3 (ITU-T X interface in ITU-T M.3010) - Customer Network Management for ATM Public Network Service which is based primarily on the IETF SNMP standards. According to the M3 specification "read only" management service (Class I of requirements) is mandatory if the service provider offers any management service.

Class I of requirements includes:

- Retrieve General UNI Protocol Stack Information
- Retrieve General ATM Level Performance Information
- Retrieve ATM Level Virtual Path/Virtual Channel (VP/VC) Link Configuration and Status Information
- Retrieve Traffic Characterisation Information
- Retrieve Event Notifications from the Public Network Provider.

Operations, Administrations and Maintenance (OAM) flows enable the tests based on the ATM Layer mechanisms. Management Information Flows 4 and 5, described in ITU-T I.610, were of special interest. These information flows (F4/F5) could be used to verify the existence of connectivity for a particular ATM connection. UNI defines F4 and F5 OAM flows on the Public UNI as End to End Loopback and UNI Loopback for respectively VPC and VCC services.

Class II of requirements is optional for the service providers. It includes addition, modification or deletion of virtual connections and subscription information in a public network. The following 6 phases based on those Class I and Class II of requirements were specified:

- Phase 1: Investigate management possibilities at each point of attachment on the user side and on the Service provider side (M3/ILMI interface).
- Phase 2: Tests of access to the Service provider management system from the User network management system like restrictions, security considerations (authentication).
- Phase 3: Tests of general monitoring functions (Class I): monitoring information on the configuration, fault and performance management on a specific user's portion of the Service provider ATM network.
- Phase 4: Tests of ATM Layer OAM End to End and Segment Loopbacks managed from the User management system for already established ATM connections (in-service measurements).
- Phase 5: Tests of advanced management functions (Class II) on a specific user's portion of the Service provider ATM network if supported: addition, modification or deletion of virtual connections and subscription information.
- Phase 6: A report of situation, experience and results of tests will be done.

Final test scenario

The JAMES didn't offer any management services in the beginning of 1996. Because of that, our final scenario, Version 3, September 1996, was based on the tests of the same management possibilities as in the initial version but applied only at the NRN ATM edge devices which participated in the TEN-34 experiments. Instead to base the tests on M3 interface - Customer Network Management for ATM Public Network Service (which needed an active role both of the NRN networks and the service provider - JAMES) we decided to test the similar functionality of M2/M3 interface based only on the NRN networks. The interface M2 (the management interface needed to manage a private ATM network) has not been standardised and in practice it has the same functionality as M3 interface.

A Management platform based on the SNMP with a special view of the NRN ATM infrastructure was included.

The management tests were more continuous than in other tests, but the following phases could be specified:

- Phase 1: Investigate the management possibilities at each point of NRN attachment to ATM overlay network.
- Phase 2: Tests of access of management system to the NRN ATM network.
- Phase 3: Tests of general monitoring functions (Class I of requirements): monitoring information on the configuration, fault and performance management.
- Phase 4: Tests of ATM Layer OAM End to End and Segment Loopbacks managed from the User management system for already established ATM connections (in-service measurements).
- Phase 5: Tests of advanced management functions: Creation of a WWW based TEN-34 management page for public access with possible manipulation with virtual connections and subscription information (Class II of requirements).
- Phase 6: A report of situation, experience and results of tests will be done. Inputs from other work packages with their experience on management issues will be taken into consideration.

Tests of access to NRN edge ATM devices

The work has started in September '96 and was continuously done as participants in tests were ready to perform the tests. Access to the edge ATM devices was realized through the public Internet service as today's ATM switches have an Ethernet access port with an Internet address which could be used for management. In cases were a firewall was applied, it was difficult to realize the transport link through the Internet and tunneled SVC ATM link was used as for UKERNA.

The SunNet Management platform from ULB/STC premises (SunNet Domain Manager version 2.3 on Solaris 2.4) was used for the access to the MIBs in eight different locations:

- NORDUnet: oslo-atm.uninett.no (128.39.2.19) read community string: public;
- INFN: miasx200.mi.infn.it (192.84.138.200) read community string: ten-34, LS1010.mi.infn.it (192.84.138.11) read community string: ten-34;
- ULB/STC: rtr02.iihe.ac.be (193.190.246.65) read community string: public;
- ACOnet: jkuatmt1.edvz.uni-linz.ac.be (140.78.2.102) read community string: TEN-34;

- SWITCH: popocatepetl.switch.ch (130.59.16.213) read community string: ten-34, castor.switch.ch (130.59.16.6) read comm. string: ten-34;
- SURFnet: atms2.cs.utwente.nl (130.89.10.230) read community string: tf-ten.
- DFN: ksatm3.rus.uni-stuttgart.de (193.196.152.2) read community string: tf-ten-nm
- UKERNA: lemon.ukerna.ac.uk (193.246.0.226) read community string: public.

Different protection levels were used for SNMP read access: public and group community strings. In the case of UKERNA the SNMP access is realized through the tunneled SVC over the ATM Overlay network,.

Tests of general monitoring functions at ATM NRN edge devices

The work started in October '96 and was continuously performed till the end of March '97. SunNet Manager Management platform used SNMPv1 "read only" management functions both for SNMPv1 and SNMPv2 agents. These tests included SNMPv1 read access and statistics collection for the different MIBs.

GARR, SWITCH, ACOnet, DFN - CISCO LightStream 1010:

- ATM-MIB
- PNNI-MIB
- CISCO-ATM-ADDR-MIB
- CISCO-ATM-CONN-MIB
- CISCO-ATM-IF-MIB
- CISCO-ATM-PHYS-MIB
- CISCO-ATM-RM-MIB
- CISCO-ATM-SWITCH-ADDR-MIB
- CISCO-ATM-TRAFFIC-MIB.

SURFnet - UB GeoSwitch 155:

- ATM-FORUM MIB
- ATM-FORUM-ADD-REG-MIB
- ATM-MIB.

GARR, UKERNA - FORE ASX200:

- ATM-FORUM-MIB
- ATM-FORUM-ADDR-REG-MIB
- FORE-SWITCH-MIB.

NORDUnet - CISCO LightStream100:

- LS100-MIB.

NORDUnet, ULB/STC, SWITCH - CISCO routers with ATM interfaces:

- SNMPv1 MIB-II.

The transport link based on tunneled SVC ATM with UKERNA had problems to receive the SNMP responses longer than 200 octets. This behaviour is due to the implementation constraints of IP over SVC through ATM tunneling.

The SNMPv2 management functions need security implementation which is not widely supported. SNMPv2 can be used on the agent side, on the manager side and or the both. On the manager side we used SNMPv1 except for elementary local tests of authentication, based on SunNet Manager realization of SNMPv2 (RFC 1446 and RFC 1447). In this realization three special files has to be created simultaneously both for the agent and the manager with control in Party Database group, Contexts Database group and Access Privileges Database group. The lack of encryption mechanisms in European versions of SNMPv2, manipulation complexity and lack of "write access" to the remote NRN ATM devices resulted in abandonment of more tests on the SNMPv2 manager side.

Through the Internet and the X-window terminal access, all participants in the tests had benefit of the SunNet Manager console system as it allows remote transparent multi-user work.

Tests of ATM Layer OAM flows

OAM flows can be applied both at the physical and at the ATM layer (ITU-TS I.610). The flows (F1, F2, F3) at the physical layer (F1, F2, F3) are dependent of the transmission system (SDH, PDH) and were not of our interest. At the ATM layer two flows: F4 and F5 are covering VP and VC level, respectively. Both flows are bidirectional and follow the same route as the user-data cells, thus constituting an in-band maintenance flow. Both ATM layer flows can either cover the entire virtual connection (End-to-End flow) or only parts of the virtual connection (Segment flow). Through the OAM flows the following groups of functions could be realized:

- Fault management, continuity check and loopback tests,
- Performance management, - System management.

Not all these functions are standardized and implemented. The loopback OAM flows are the first being standardized and implemented and therefore they were of our primarily interest. The loopback tests enable the verification of ATM Layer connectivity existence for a particular connection. For F4 flow VPI corresponds to tested VP and VCI is constant, always set to 4. For F5 flow VPI and VCI correspond to tested VC. The mechanism consists of sending out the loopback cells and activating timers. If the originator receives back the looped cells in the interval of 5 seconds it is assumed that the connectivity exists. In practice the OAM flows could be activated through a customer access UNI or through the TMN (Telecommunications Management Network). The tests we performed overthe Overlay ATM network were End to End Loopback tests activated from the UNI at CISCO LightStream 100 and CISCO LightStream1010. Beside that, the CISCO-ATM-OAM-MIB offers a possibility to activate the OAM flows through the SNMP based management interface. This second functionality was not tested as it requires "write" i.e. "set" possibility at the NRN ATM switches while we had "read only" access.

An example of OAM tests realized in March '97 is described here. These End-to-End OAM F5 flow tests were performed from ACOnet to SWITCH, DFN and GARR, all with CISCO LS1010s at both ends. The obtained Round Trip Time (RTT) were:

1. ACOnet-SWITCH: Minimum RTT=12 msec, Average RTT=12 msec, Maximum RTT=12 msec
2. ACOnet-DFN: Minimum RTT=32 msec, Average RTT=34 msec, Maximum RTT=36 msec
3. ACOnet-GARR: Minimum RTT=12 msec, Average RTT=14 msec, Maximum RTT=16 msec.

In the tests the sequences of 5 OAM cells with 53 octets were sent. The results show that RTT using OAM loopback cells has rather constant values (the standard deviation in our tests is between 0 and 2). This behavior is due to the flows on the pure cell level. For the OAM loopback flows the switches have to perform very simple checking: on loopback type (End-to-End/Segment), on indicator (forwarding flow/backwarding flow), on correlation tag (unique id. flow number) and on connection end-point location identifier. The last parameter is used in the loopback tests variant called Loopback Test Using Loopback Location Identifier as in the case of CISCO implementation. The Loopback Location Identifier is not standardized and at CISCO it is an ATM switch address prefix. The experience with OAM flows shows that the mechanism even in its early implementation phase is very promising manner to learn about ATM layer behavior throughout the large ATM networks.

5.6.9 Test related problems and general suggestions

The test scenario has been changed from the tests on public to private ATM segments due to the lack of the general management services offered by PNOs (JAMES). The modifications of the test scenario assumed the existence of the similar functionality on the edge NRN devices and on the public segments which correspond to the specific user.

Common ATM based MIBs were not uniform: not all had the same groups implemented what made presentation and comparison difficult. For example FORE ASX200 supports the following groups in ATM-FORUM MIB: atmfAtm tatsTable, atmfVccTable and atmVpcTable, while UB GeoSwitch 155 supports only atmfAtmStatsTable and CISCO LS1010 supports none of them.

Other specific functions like "OAM segment Loopback" needed to be tested together with JAMES. In this phase JAMES didn't show an interest for it.

5.6.10 Relevance for service and migration suggestions

SNMP based management platforms on the user premises could be used for the M3 interface (Customer Network Management for ATM Public Network Service). As SNMPv2 is not completely implemented the problem of access control is still unsolved what can be problem for the Class II of requirements. The CMIP based management platforms with X.user interface could offer more secure functionality but the process is still in the standardisation phase. OAM flows which allow the transfer of management information between different management protocols can be successfully used in End-to-End as well in Segment test scenarios. Web based ATM management interface could offer different levels of security and user-friendly management functions what can be used to overcome the problems of security with SNMP.

The other problem concerned the delay in realisation of the Web based management of TEN-34 ATM network caused by the lack of the uniformed MIBs at VP/VC level for statistics presentation and by the lack of standardised tools for the Web based management.

5.6.11 Further studies

The future realisation of X.user interface and possible co-operation with JAMES will be studied. The relevant standards concern the new releases of existing ITU-T standards like M.3020 (TMN Interface Specification Methodology) and M.3100 (Generic Network Information Model) and new in-progress ITU-T standards like M.3203 (Customer controlled service management) and M.3205 (B-ISDN management) will be studied.

5.6.12 Bibliography and References

[1] A. Guillen, Z. Cekro: Belgian ATM platform, Backup Application, ULB/STC participation, June 1996

[2] ATM Forum Specification: Customer Network Management for ATM Public Network Service (M3 Specification), 1996

[3] ATM Forum Specification: UNI v. 3.1, 1995

[4] ATM Forum Specification: UNI v. 4.0, 1996

[5] ATM Forum Specification: ILMI v. 4.0, 1996

[6] ATM Forum Specification: Introduction to ATM Forum Performance Benchmarking Specifications, 1996

[7] ITU-T, I.610, Integrated Services Digital Network (ISDN), Maintenance Principles, B-ISDN Operation and Maintenance Principles and Functions, November 1995

[8] ITU-T, I.751, Integrated Services Digital Network (ISDN), B-ISDN Equipment Aspects, Asynchronous Transfer Mode, Management of Network Element View, March 1996

[9] ITU-T Recommendation I.356, B-ISDN ATM layer cell transfer performance, 1993

[10] CCITT Recommendation M.20, Maintenance philosophy for telecommunications networks, 1992

[11] CCITT Recommendation M.3010, Principles for a telecommunications management network, 1992

[12] IETF RFC 1446, J.Galvin, K.McCloghrie, Security Protocols for version 2 of the Simple Network Management Protocol (SNMPv2), 1993

[13] IETF RFC 1447, K.McCloghrie, J. Galvin, Party MIB for version 2 of the Simple Network Management Protocol (SNMPv2), 1993

[14] IETF RFC 1695, M. Ahmed, K.Tesink, Definitions of Managed Object for ATM Management Version 8. using SMIv2.

5.7 CDV over concatenated ATM networks

5.7.1 Experiment Leaders

Victor Reijs, SURFnet, NL
P. F. Chimento, University Twente, NL

5.7.2 Participants and Equipment

There were 4 organizations that participated in the measurements: The University of Twente, KPN Research in Leidschendam, the Netherlands, the University of Stuttgart and Deutsche Telekom in Kˆln. Each of these organizations captured the traces with their ATM analyzer equipment. The people involved were Robert Stoy at the University of Stuttgart, Dirk Hetzer at DT Berkom in Berlin, Mr. Schurillis from DT in Kˆln, Harrie van de Vlag at KPN Research and Edward Meewis and Phil Chimento at the University of Twente.

The measurements were made on 5 different days: 2 days in December 1996 and 3 days in mid-January 1997. The measurements were all made in the mid- to late afternoon on these days, and it is estimated by the participants that there was very little other traffic in the network on those days at those times.

	U Twente	RU Stuttgart	DT BERKOM	KPN
Switch(es)	UB Networks GeoSwitch	LS 1010	Siemens EWSX	GDC and AT&.T GV2000
Monitor	HP 5200A	W&G DA30c	HP 75000 and HP 5200	HP 75000
VPI/VCI available	No restrictions	15 VPIs, 200 VCIs	No VPI 0	no restrictions
Line Speeds	155 Mb/s	155 Mb/s	34 Mb/s + 155 Mb/s	34 Mb/s+155 Mb/s
Time Stamp resolution	100 ns	10 microsec	100 ns	100 ns
Trace buffer	114 Kcell	16 MB	> .110 Kcell	> .110 Kcell

Table 1: Equipment and capabilities at test sites

Table 1 shows the equipment used by the various participants. The major difference in the measurement equipment was due to the time-stamp resolution. All the HP analyzers have a resolution of 100 nanoseconds, while the W&G analyzer has a resolution of 10 microseconds. This did not cause any problems, but you will notice that the results for the measurements at Stuttgart are less accurate than those for the other sites.

5.7.3 Configurations

KPN began its participation in January 1997, and therefore, at that time, the path between the University of Twente and the University of Stuttgart changed. Figure 2 shows the original experimental setup between the University of Twente and the University of Stuttgart for the tests that were run in December 1996. In the experiments run in this configuration, we have cell stream traces from University of Twente, DT and University of Stuttgart.

Figure 1 is the configuration after KPN also began to measure. This considerably lengthened the path that was traversed by the cell streams that were sent during the experiments.

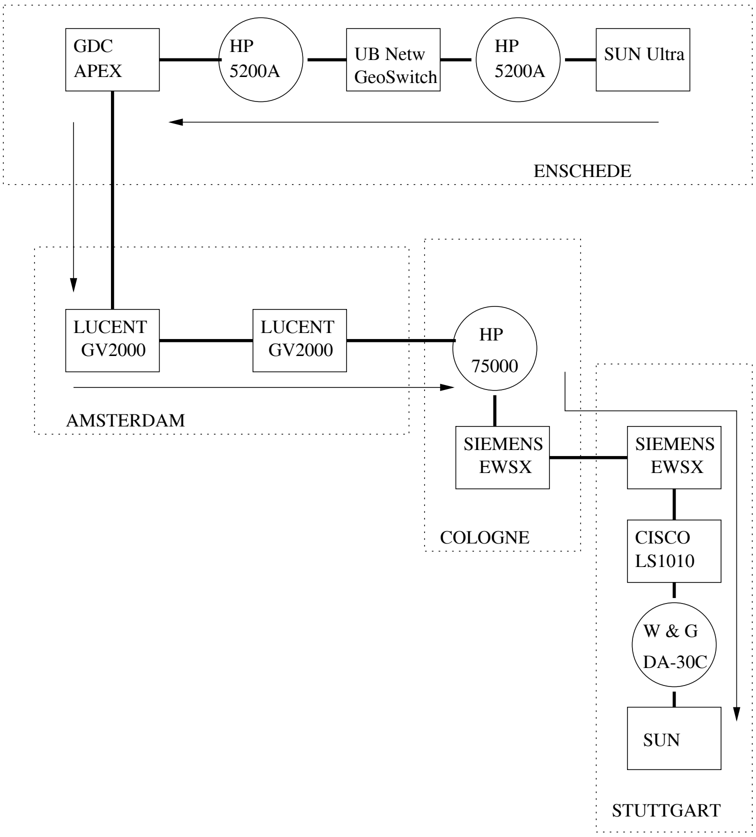


Figure 1: Short Path between Enschede and Stuttgart

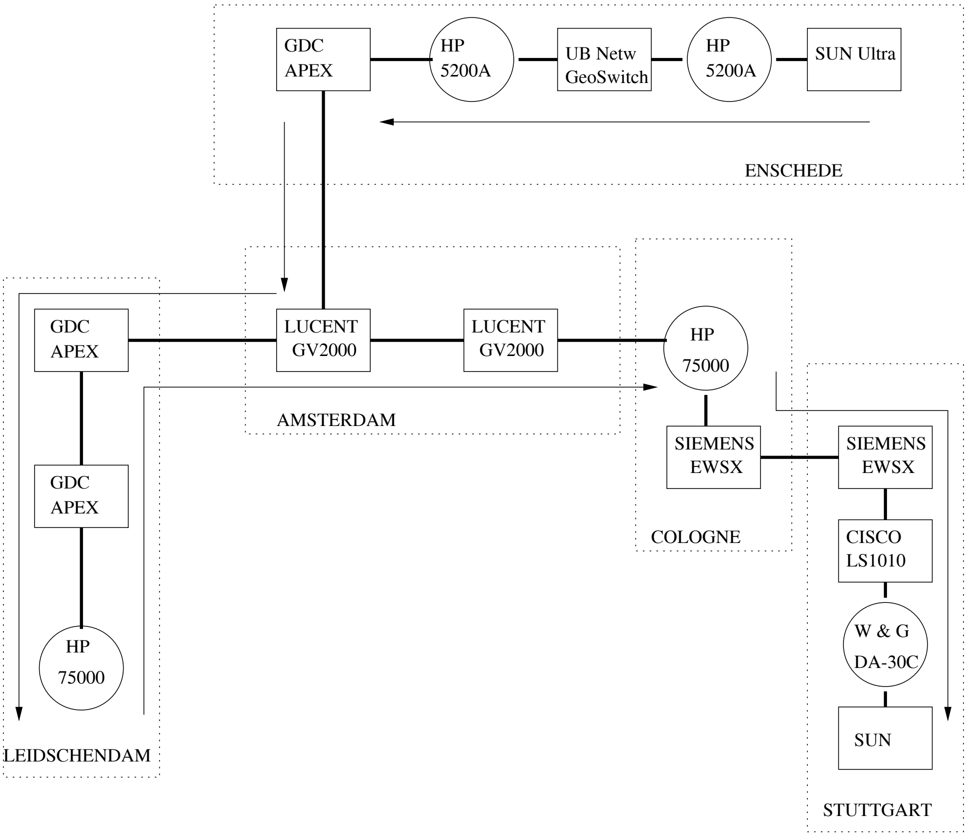


Figure 2: Long Path between Enschede and Stuttgart

The following list gives a short description of the measurement points that were active during one or another part of the tests. In the rest of the document, we will use the abbreviations for the measurement points given in this list.

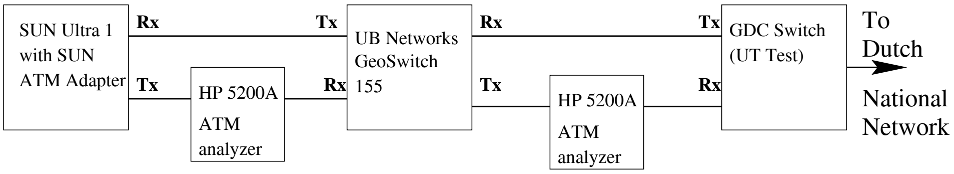
UT1 This measurement point was placed at the tailgate of the SUN Ultra which was generating the cell stream. During the measurements in December 1996, this was accomplished by bringing the cell stream into the HP 5200A and retransmitting it from the HP to the UB Networks Geoswitch. Starting in January 1997, we had optical splitters which brought the cell stream to the HP analyzer, but what reached the Geoswitch was transmitted directly from the SUN Ultra. Figure 3 shows the details of the configuration was used for the experiments performed in January.

UT2 This measurement point was only active during the January 1997 measurements and was placed directly after the UB Networks Geoswitch. It was also fed via an optical splitter leading to the HP 5200A. During January, there was no possible interference in the behaviour of the cell stream from either of the University of Twente measurement points.

KPN This measurement point was an HP 75000 located in Leidschendam at KPN Research, one of the partners in JAMES. In effect, this measurement point was placed in the middle of the Dutch national network, and captured the cell stream before it got to the international part of the path. The KPN measurement point was active only during the measurements taken in January 1997.

DT This measurement point was an HP analyzer at K^ln at Deutsche Telekom in Germany, also one of the partners in JAMES. This point was located directly in the path of the international link between the Netherlands and Germany which is a part of the JAMES network. Except for the very first trial on 20-12-96, this measurement point was active for all the experiments.

RUS This measurement point was placed just after the final switch in Stuttgart, between the last switch (a Cisco LS 1010) and the receiving SUN system.



Detailed picture of the configuration at the University of Twente

Figure 3: Local Configuration as UT

During the measurements over the short path, there were at most 3 measurement points active: UT1, DT and RUS. We give a summary of the number and type of switches in the path:

UT1->DT

There were 4 ATM switches in this path:

1. UB Networks GeoSwitch
2. GDC Apex Switch
3. 2 Lucent Technologies Globeview 2000 switches

DT->RUS

There were 3 ATM switches in this path:

1. 2 Siemens EWSX ATM switches
2. Cisco Lightstream 1010

The measurements over the long path had as many as 5 measurement points active: UT1, UT2, KPN, DT and RUS. The numbers of switches on the parts of these paths were as follows:

UT1->UT2

There was only one switch on this part of the path: the UB Networks GeoSwitch.

UT2->KPN

There were 4 ATM switches on this part of the path:
1. 3 GDC Apex ATM switches
2. Lucent Technologies Globeview 2000 switch

KPN->DT

There were 4 ATM switches on this part of the path:
1. 2 GDC Apex ATM switches
2. 2 Lucent Technologies Globeview 2000 switches

DT->RUS

There were 3 ATM switches in this path:
1. 2 Siemens EWSX ATM switches
2. Cisco Lightstream 1010

The official data on the VPC between the Netherlands and Germany which we used is as follows:

Name of the connection: Es-p36-Stgt/Atm1
KPN/SURFnet Switch in NL : Utw-test
Physical Port/VPI/VCI: ST1A/6/*
Destination : RU Stuttgart
Bandwidth: 4750 cells/s (1.824 Mbit/s)
Burstsize: 1
Start: 96/12/02/09/00
End: 97/01/31/23/59

This connection was a permanent VPC which was established by KPN at the request of the experimenters.

Traces Gathered

The traces that were actually captured during the experiments are summarized below. These traces will all be available via the WWW at the URL: http://wwwtios.cs.utwente.nl/chimento/tf10exp/tracelist-1.html. Table 2 gives a summary of the traces that exist from these experiments. In total, there are 62 traces collected at the 5 measurement points. There were 17 trials that succeeded at least partially and one where all the traces were lost because of measurement equipment failure, and one trial which did not succeed because someone inadvertently re-IPLd a switch at the University of Twente.

Number	Trial	Date	Time	UT1	UT2	KPN	DT	RUS
	Short Path between Enschede and Stuttgart							
1	1.1 Mb/s	20-12-96	14:53	Yes	No	No	No	Yes
2	1.1 Mb/s	23-12-96	15:08	Yes	No	No	Yes	Yes
3	1.1 Mb/s	23-12-96	15:20	Yes	No	No	Yes	Yes
4	1.6 Mb/s	23-12-96	15:38	Yes	No	No	Yes	Yes
5	1.6 Mb/s	23-12-96	15:54	Yes	No	No	Yes	Yes
6	1.6 Mb/s	23-12-96	16:07	Yes	No	No	Yes	Yes
	Long Path between Enschede and Stuttgart							
7	1.6 Mb/s	14-1-97	15:00	Lost	Yes	Lost	Yes	Yes
8	1.6 Mb/s	14-1-97	15:13	Yes	Lost	Lost	Yes	Yes
9	1.6 Mb/s	14-1-97	16:40	Yes	Yes	Yes	Yes	Yes
10	1.1 Mb/s	14-1-97	16:50	Lost	Lost	Yes	Yes	Lost
11	1.1 Mb/s	16-1-97	16:19	Yes	Yes	Yes	Yes	Yes
12	1.1 Mb/s	16-1-97	16:38	Yes	Yes	Yes	Yes	Yes
13	1.1 Mb/s	16-1-97	16:58	Yes	Yes	Yes	Yes	Yes
14	VBR	17-1-97	14:26	Yes	Yes	Yes	Yes	Yes
15	VBR	17-1-97	15:58	Yes	Yes	Yes	Yes	Lost
16	0.5 Mb/s	17-1-97	16:39	Lost	Lost	Yes	Yes	Yes
17	0.5 Mb/s	17-1-97	17:07	Yes	Yes	Yes	Yes	Yes

Table 2: Traces taken during the CDVT/BT experiments

The trials were coordinated via telephone and e-mail. At the University of Twente, we coordinated with KPN and University of Stuttgart and Stuttgart contacted DT.

Trial number 7 in the list contains an anomaly (an extraordinarily long interarrival time) which was probably caused by something in the SUN Ultra generating the cell stream. The traces from this trial will not be used in this report.

We have not yet completed the processing of the two VBR experiments, and so those results are also not included in this report.

5.7.4 Cell-stream generation

Cell stream generation was done by a program developed at the University of Twente, based on a sample program provided by Sun Microsystems with the software for their Sun ATM adapter. We found that below 2 Mb/s, the Sun Ultra, as a CBR stream generator was very stable, producing cell streams with a very narrow distribution of cell interarrival times. This is quantified somewhat below in Section *Changes in Cell Interarrival Times*, but a more complete characterization of the Sun Ultra as a CBR stream generator is given in [1].

5.7.5 Observations

Changes in cell interarrival times

IAT Spread

Table 3 summarizes the spread of cell interarrival times as measured during these experiments. In this table, we have grouped the results by experiment (i.e. by the speed at which the cell stream was transmitted) and by the path that the cell stream took. The numbers in Table 3 are IAT spread defined as for each trace, and are reported in units of microseconds.

		Measurement Point				
Trial No.	Experiment	UT1	UT2	KPN	DT	RUS
1*	1.1 Mb/s	3.3	-	-	-	70+-9
2*	1.1 Mb/s	3.3	-	-	53.9	100+-9
3*	1.1 Mb/s	3.3	-	-	62.0	100+-9
10	1.1 Mb/s	-	-	65.3	93.7	-
11	1.1 Mb/s	3.3	8.8	65.3	91.0	130+-9
12	1.1 Mb/s	3.3	8.8	65.3	110.9	130+-9
13	1.1 Mb/s	3.3	8.8	64.8	113.2	130+-9
4*	1.6 Mb/s	3.3	-	-	62.1	100+-9
5*	1.6 Mb/s	3.3	-	-	53.9	100+-9
6*	1.6 Mb/s	3.3	-	-	51.2	100+-9
8	1.6 Mb/s	3.3	-	-	115.3	130+-9
9	1.6 Mb/s	3.3	9.3	65.3	99.6	130+-9
16	500 kb/s	-	-	76.7	102.2	140+-9
17	500 kb/s	3.3	9.3	76.3	91.0	140+-9

Table 3: Spread of cell interarrival times. The table contains the difference between the largest and smallest cell interarrival times in the trace indicated. Times are all in microseconds.

From Table 3 one can see that at the source, (i.e. directly after the cell stream exits the SUN which generates it) the maximum difference in cell interarrival times is a bit more than one cell time (which is 2.8 microseconds at 155 Mb/s). This difference is at least in part due to the slotted nature of ATM, and probably in part due to the cell stream generating mechanism. However, across speeds and experiment trials, the maximum difference at this point of measurement does not change, indicating a stable source.

One important point to notice is the difference in the IAT spread between the short path and the long path between Enschede and Stuttgart. On the short path (Trials 1-6) the number of switches between UT1 and DT is 4 and the increase in the IAT spread is roughly 50 microseconds. On the long path, the number of switches between UT1 and KPN is 5 and the increase in IAT spread is a bit more, about 60 microseconds (through different switches).

We can see similar differences between KPN and DT on the long path, where the increase in IAT spread is roughly 40-50 microseconds for 4 switches between
these two points, for the 1.1 Mb/s and 1.6 Mb/s experiments. For the 500 kb/s experiment, the increase in IAT spread is somewhat less. Between the DT and RUS measurement points, the IAT spread increases vary more with the path used. For the short path, the increase is 40-50 microseconds, but for the long path, the increase is 20-30 microseconds, except for the 500 kb/s experiment where the increase is about 40 microseconds again. The number of switches between these two points was 3.

5.7.6 Distribution of differences from the mean

Two of the points discussed at the TF-TEN meeting in Z¸rich in March 1997 were the dependency of successive interarrival times, and whether the distribution of the differences between the cell interarrival times and the mean interarrival time were normally distributed. We briefly explored these questions after the meeting and results in Table 4 and the empirical distribution functions in Appendices A, B and C show the results of this analysis. In this section, the difference from the mean is defined as follows: Take the sample mean of the trace, and subtract the sample mean from each cell interarrival time computed from the trace. It is the distribution of these differences which is being analyzed here.

The issue of dependency is something that we will continue to investigate as a part of this project. The dependency between successive interarrival times is somewhat hard to determine since it is not the case that a shift of one cell in the stream automatically causes successive interarrival times to be smaller then larger than the mean. In examples that we have run on the traces, there are runs of a number of interarrival times which are below the mean followed by one that is above the mean. In all cases, the mean difference between interarrival times and the average for each trace is 0 to within 6 decimal places, though in fact, the mean difference shows some slight positive or negative tendency. It is clear that we need to do more analysis on this point.

There are strong indications that the distribution of differences between the mean of each trace and the interarrival times is not normal. The empirical cumulative distribution functions (ecdfs) for each of the traces considered in this report is given in the appendices. The shapes of these functions do not appear to be either symmetric or normal; however, there are more tests which can be applied to determine what the distributions 'fit'.

		Measurement Point									
Trial No.	Experim.	UT1		UT2		KPN		DT		RUS	
		+	-	+	-	+	-	+	-	+	-
1*	1.1 Mb/s	.673	.327	x	x	x	x	x	x	.382	.618
2*	1.1 Mb/s	.673	.327	x	x	x	x	.485	.515	.384	.616
3*	1.1 Mb/s	.673	.327	x	x	x	x	.486	.514	.384	.616
10	1.1 Mb/s	x	x	x	x	.421	.579	.487	.513	x	x
11	1.1 Mb/s	.673	.327	.653	.347	.478	.522	.504	.496	.392	.608
12	1.1 Mb/s	.673	.327	.653	.347	.423	.577	.490	.510	.395	.605
13	1.1 Mb/s	.673	.327	.653	.347	.405	.595	.485	.515	.395	.605
4*	1.6 Mb/s	.782	.218	x	x	x	x	.323	.677	.548	.452
5*	1.6 Mb/s	.782	.218	x	x	x	x	.323	.677	.548	.452
6*	1.6 Mb/s	.782	.218	x	x	x	x	.323	.677	.548	.452
8	1.6 Mb/s	.782	.218	x	x	x	x	.400	.600	.538	.462
9	1.6 Mb/s	.782	.218	.746	.254	.330	.670	.400	.600	.538	.462
16	500 kb/s	x	x	x	x	.369	.631	.508	.492	.619	.381
17	500 kb/s	.347	.653	.374	.626	.369	.631	.510	.490	.620	.380

Table 4: Distribution of cell interarrival times above the mean (+) and below the mean () of each trace. The '*' denotes that the cell stream traversed the short path.

As further evidence that the distribution of differences from the mean is not normal, Table 4 shows the fraction of samples above the mean (in the columns marked '+') and below the mean (in the columns marked '-') for each measurement point. The interesting point about this table is that these fractions change (and in fact, reverse themselves) as the cell stream travels through the network. These fractions also appear to be dependent on the speed at which the cell stream is sent. For the 1.1 Mb/s experiment, the cell stream as generated has about of the samples above the mean and below the mean. By the time the cell stream arrives at Stuttgart, the proportions are reversed. For the 1.6 Mb/s experiment, about 80% of the interarrival times are above the mean and 20% below at first, and when the cell stream arrives in Stuttgart the proportions are about 55% above and 45% below the mean. Clearly this phenomenon needs more analysis.

5.7.7 Changes in GCRA parameters

In order to look at the effect of the ATM network on the traffic descriptors, we ran the traces through a program to simulate the operation of a GCRA (continuous state leaky bucket) as defined in the ATM Forum and ITU documents. This program is able to produce, for a given inter-cell time, the worst case leaky bucket 'depth' required for the cell stream to pass through the GCRA with no cells being flagged in violation of the GCRA. This produces, in some sense, the 'smallest' set of GCRA parameters needed for no violations. By varying the inter-cell time, we were able to produce zero-violation curves, which show the behaviour of the cell stream over a range of GCRA parameters. For the 500 kb/s experiment, these are given in Figures 30 and 31 For the 1.1 Mb/s experiment, they are Figures 11, 12, 13, 14, 15, 16, 17. For the 1.6 Mb/s experiments, these are Figures 23, 24, 25, 26, 27. Note that the L parameter (y-axis, which is the CDVT) is on a log scale.

First, a few notes about the general shape of the zero violation curves: The only interesting part of the curve is the left side. That is the side where the inter-cell time is less than or equal to the nominal inter-cell time (i.e. the rate at which the stream is supposed to be policed). The sharp rise in the L parameter (i.e. the CDVT) at about the nominal inter-cell time is explained as follows: at that point, the bulk of the cells are arriving faster than the policed rate. Since that is the case, the CDVT must increase rapidly in order to account for what are increasingly long trains of cells which arrive at the leaky bucket before it is empty.

The sharp drop-off at the left tail of these curves is due to the fact that at that point, the minimum inter-cell time is reached and no cells arrive earlier than the policed inter-cell time and thus the stream has a CDVT of 0. These phenomena have been observed and explained in more detail by [2].

			Measurement Points				
		policed	CDVT (ms)				
Trial No.	Experiment	IAT (ms)	UT1	UT2	KPN	DT	RUS
1*	1.1 Mb/s	.3810	.0021	-	-	-	.0630
2*	1.1 Mb/s	.3810	.0021	-	-	.0328	.0630
3*	1.1 Mb/s	.3810	.0021	-	-	.0334	.0630
10	1.1 Mb/s	.3810	-	-	.0380	.0560	-
11	1.1 Mb/s	.3810	.0021	.0054	.0376	.0559	.0860
12	1.1 Mb/s	.3810	.0021	.0054	.0412	.0611	.0860
13	1.1 Mb/s	.3810	.0021	.0054	.0349	.0619	.0860
4*	1.6 Mb/s	.2540	.0025	-	-	.0304	.0700
5*	1.6 Mb/s	.2540	.0025	-	-	.0304	.0680
6*	1.6 Mb/s	.2540	.0025	-	-	.0338	.0680
8	1.6 Mb/s	.2540	.0025	-	-	.0603	.0900
9	1.6 Mb/s	.2540	.0025	.0054	.0408	.0563	.0860
16	500 kb/s	.7625	-	-	.0378	.0664	.0875
17	500 kb/s	.7625	.0030	.0058	.0400	.0629	.0850

Table 5: Leaky bucket parameters computed for the various traces. The '*' denotes that the cell stream traversed the short path.

These graphs give an overall view of how the CDVT changes over the range of policed rates. More specific information can be gotten from Table 5. Here, for each of the experiments, we look at one specific point in the zero-violation curve, the nominal inter-cell time corresponding to the rate of the CBR cell stream. Here we can compare the behaviour of the cell streams with respect to their traffic descriptors on the long and short paths.

Once again, the first thing to notice from Table 5 is that the CDVT computed for the traces observed at the DT measurement point on the short path and that computed for the traces observed at the KPN measurement point on the long path are similar. In the first case, there are 4 switches between the source and the measurement point and in the second case there are 5 switches. The same holds true for RUS on the short path (7 switches) and DT on the long path (9 switches). These similarities give some indication that the increase in CDVT is in fact introduced by the switches.

From Table 5 we can examine the differences in CDVT between the different measurement points. Because there are so few trials of each experiment, we combine all the experiments on the long path to determine the differences in CDVT for the long path and likewise combine all the measurements on the short path to determine the differences in CDVT for that path. The average difference in CDVT between measurement points for each path is as follows:

1. The long path:

UT1-UT2	1 switch, 3.12 microseconds, 3.12 microseconds per switch
UT2-KPN	4 switches, 33.42 microseconds, 8.355 microseconds per switch
KPN-DT	4 switches, 21.46 microseconds, 5.36 microseconds per switch
DT-RUS	3 switches, 25.96 microseconds, 8.65 microseconds per switch

2. The short path:

UT1-DT	4 switches, 29.82 microseconds, 7.455 microseconds per switch
DT-RUS	3 switches, 34.24 microseconds, 11.41 microseconds per switch

Because we have so few trials in total, it is not known yet whether the differences in the estimate of CDVT added per switch is statistically significant or not.

5.7.8 Conclusions

After observing the behavior of the cell streams in the experiments, we can put forward a number of tentative conclusions:

1. Raw measures, such as the cell stream spread, seem to depend only on the path length and not on the speed of low speed (i.e.) cell streams.
2. The distributions of the inter-cell arrival times are characterized by exactly the same means (10s of nanoseconds) with variance increasing with path length. Outliers appear on both upper and lower ends of the distributions.
3. The GCRA analysis shows:
    1. Significant increase in the CDVT for a fixed PCR for the CBR streams
    2. It would appear that the safest course is to specify a 'loose' GCRA to describe a CBR cell stream.

However, there may be a number of causes that contribute to this CDVT behaviour:

1. Influence of the measurement instruments: Though there are indications from the data that this was not significant, (such as the similarity of differences in CDVT for similar numbers of switches) we would like to try to quantify this, or at least to produce a set of measurements with all instruments using splitters so that this is not a factor.
2. Influence of changes in physical layer on the increase in CDVT: In the experiments performed so far, the physical layer changed at least 3 times along the path: from STM-1 to SONET, from SONET to PDH, and from PDH to STM-1 again. In continuing work we will try to quantify the effect that these changes have on the cell stream.
3. Influence of switches themselves: We will continue to try to isolate this effect and to quantify it better.

5.7.9 What remains to be done

This study has been an initial rough study of the changes in CDVT/BT in ATM networks. There is certainly additional analysis to be done on the cell streams collected so far, in terms of a more 'dynamic analysis', that is, the analysis of patterns within the behavior of a cell stream and the analysis of how such patterns change across cell streams. One example of this is to look at the 'runs' of interarrival times that are longer than average or shorter than average. Another example, is to look at the time between cells at a specific position in the cell stream, at different points in the path.

It is clear that more experiments and more trials of each experiment need to be done before any solid conclusions can be drawn. Whether apparent differences in behavior of the cell streams on different paths and at different speeds is statistically significant or not is not known at this point. Further experimentation should help to determine this.

There were additional goals set out at the beginning of the project which were not realized because of limitations of time and equipment and participation from other partners. Some of these are:

1. Study CBR at higher speeds: We would like to do this in order to learn whether the changes in the CDVT/BT of the cell stream change more when the CBR stream approaches the speed of the link.
2. Study true VBR connections in JAMES: The purpose of this would be to see whether there is a difference in the behaviour of the network with respect to VBR connections, and how the SCR leaky bucket (in addition to the PCR leaky bucket) is changed.
3. Add and Control background traffic: The purpose would be to study the effect of different levels of background traffic on the (VBR and CBR) cell streams. We expect it to have an effect, but the purpose here would be to try to quantify it.

5.7.10 References

[1] C.E.P.J. Meewis. Changes in traffic characteristics and traffic descriptors in concatenated ATM networks. Master's thesis, Universiteit Twente, April 1997.

[2] A.M.R. Slingerland. A study of CDV tolerance in the specification of a source traffic descriptor for ATM systems. Master's thesis, Universiteit Twente, March 1996.

Experiment with 1 Mb/s CBR

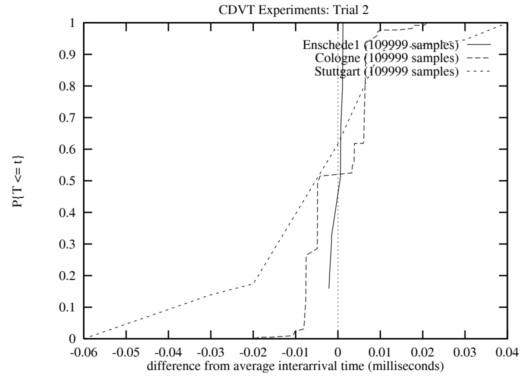


Figure: Empirical CDF: Trial 2: 1 Mb/s

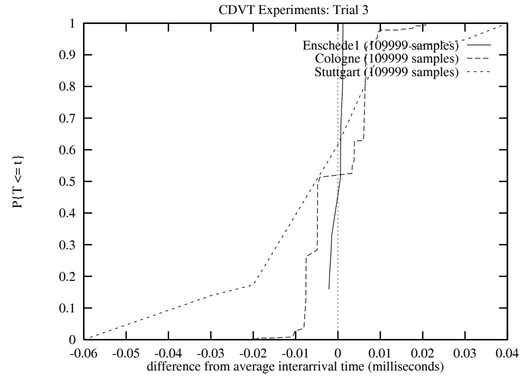


Figure: Empirical CDF: Trial 3: 1 Mb/s

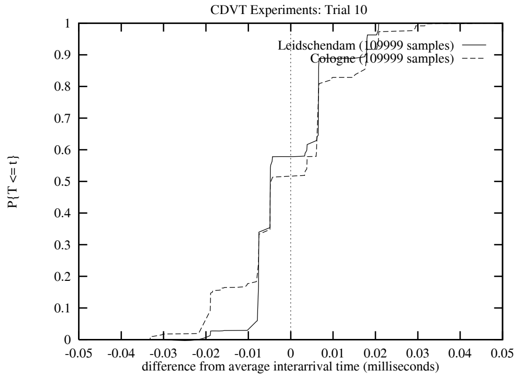


Figure: Empirical CDF: Trial 10: 1.5 Mb/s

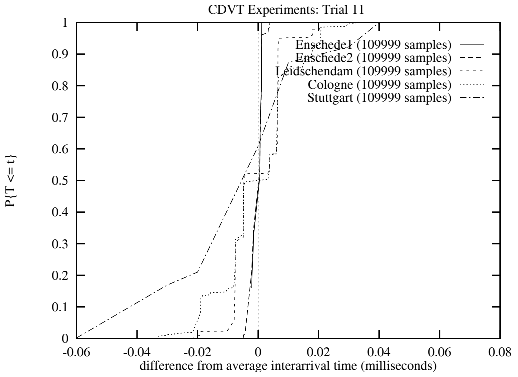


Figure: Empirical CDF: Trial 11: 1 Mb/s

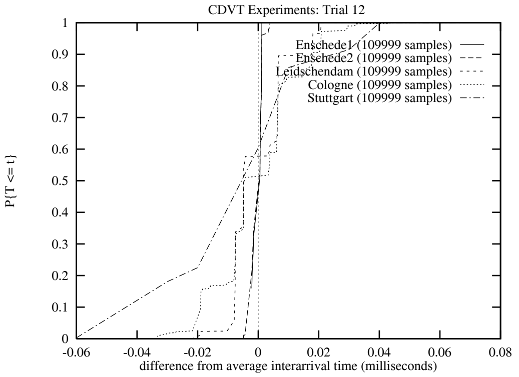


Figure: Empirical CDF: Trial 12: 1 Mb/s

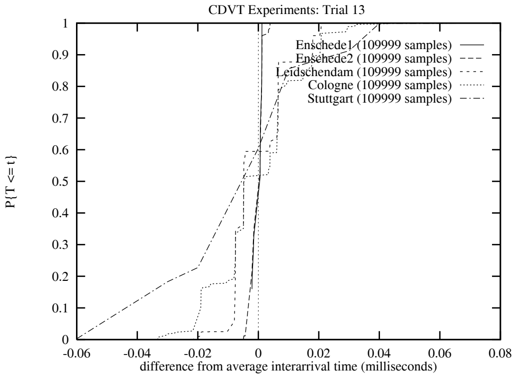


Figure: Empirical CDF: Trial 13: 1 Mb/s

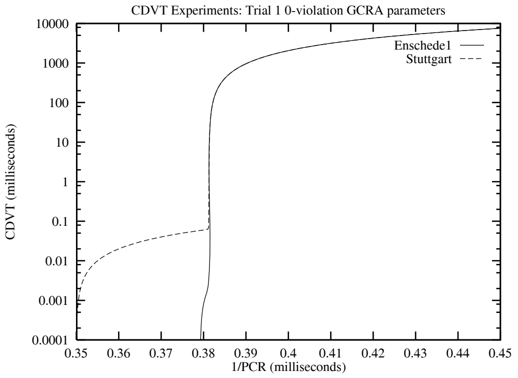


Figure: GCRA Analysis: Trial 1: 1 Mb/s

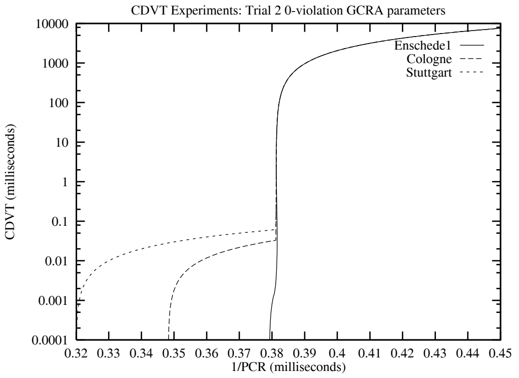


Figure: GCRA Analysis: Trial 2: 1 Mb/s

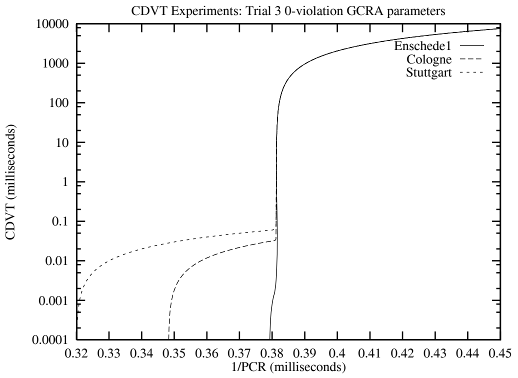


Figure: GCRA Analysis: Trial 3: 1 Mb/s

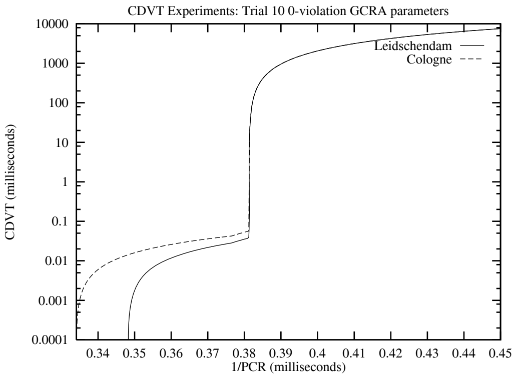


Figure: GCRA Analysis: Trial 10: 1 Mb/s

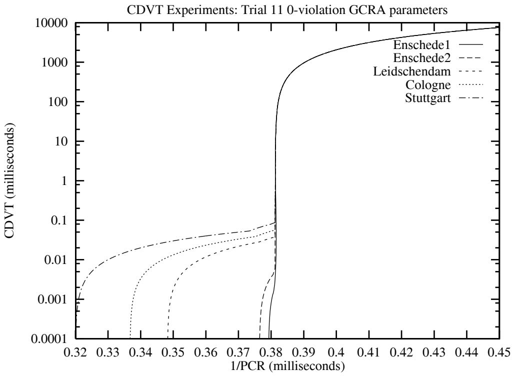


Figure: GCRA Analysis: Trial 11: 1 Mb/s

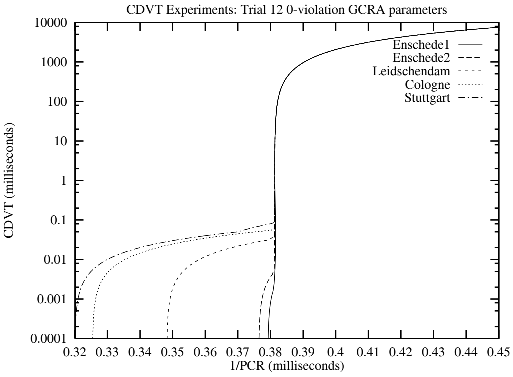


Figure: GCRA Analysis: Trial 12: 1 Mb/s

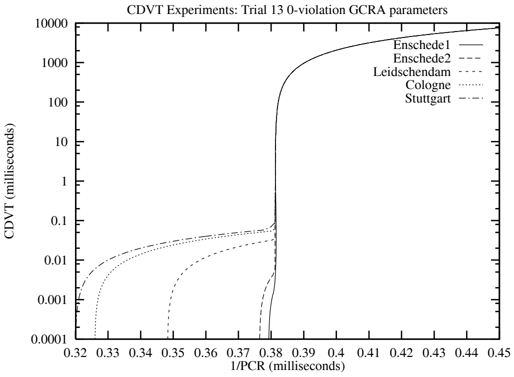


Figure: GCRA Analysis: Trial 13: 1 Mb/s

Experiment with 1.5 Mb/s CBR

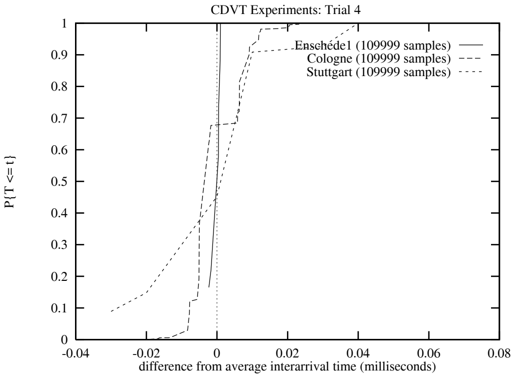


Figure: Empirical CDF: Trial 4: 1.5 Mb/s

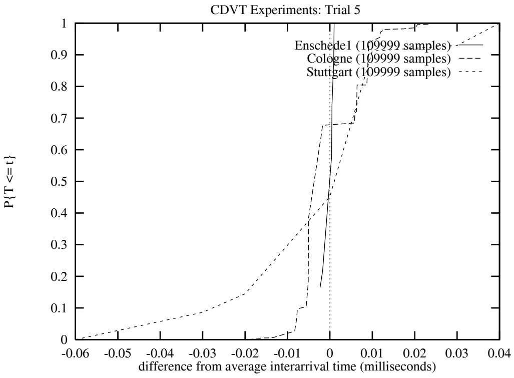


Figure: Empirical CDF: Trial 5: 1.5 Mb/s

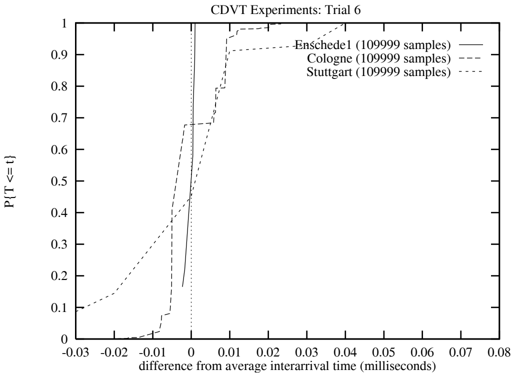


Figure: Empirical CDF: Trial 6: 1.5 Mb/s

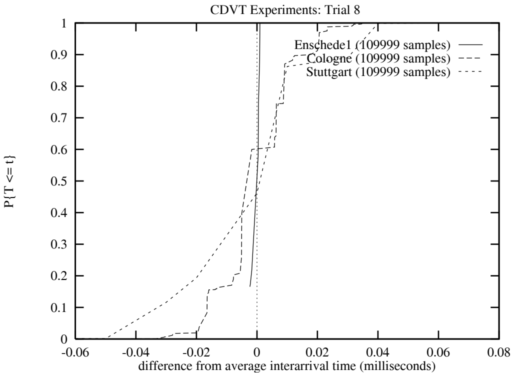


Figure: Empirical CDF: Trial 8: 1.5 Mb/s

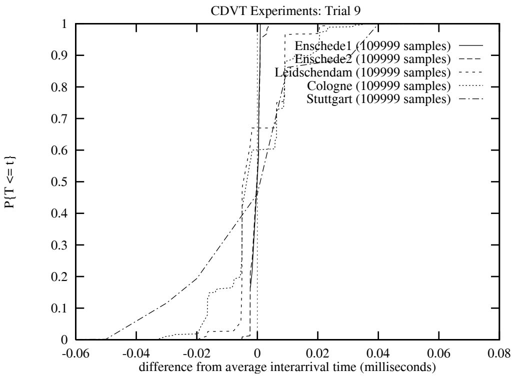


Figure: Empirical CDF: Trial 9: 1.5 Mb/s

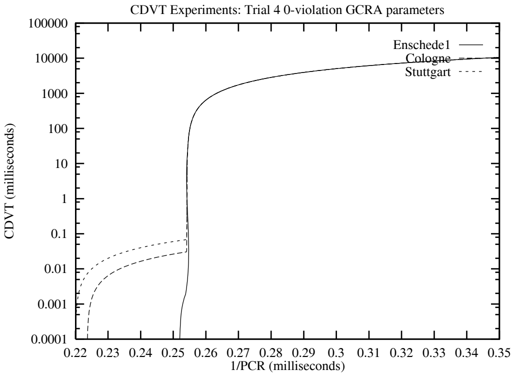


Figure: GCRA Analysis: Trial 4: 1 Mb/s

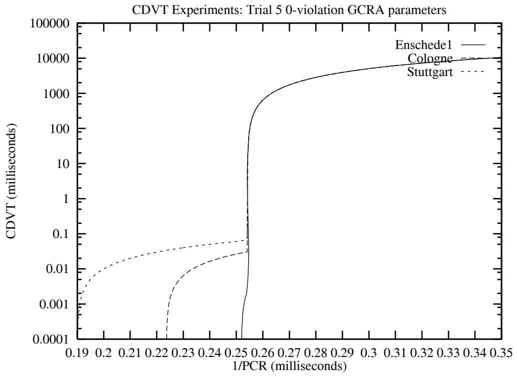


Figure: GCRA Analysis: Trial 5: 1 Mb/s

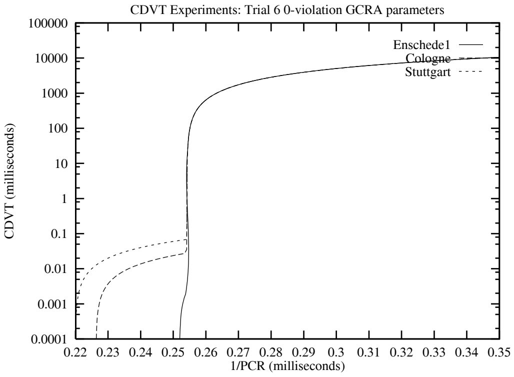


Figure: GCRA Analysis: Trial 6: 1 Mb/s

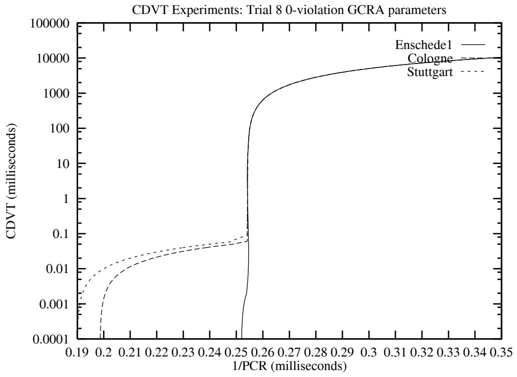


Figure: GCRA Analysis: Trial 8: 1 Mb/s

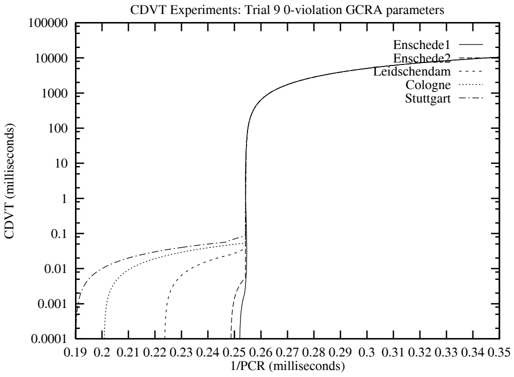


Figure: GCRA Analysis: Trial 9: 1 Mb/s

Experiment with 500 kb/s CBR

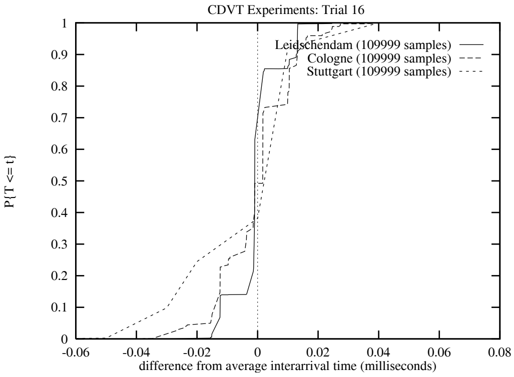


Figure: Empirical CDF: Trial 16: 500 kb/s

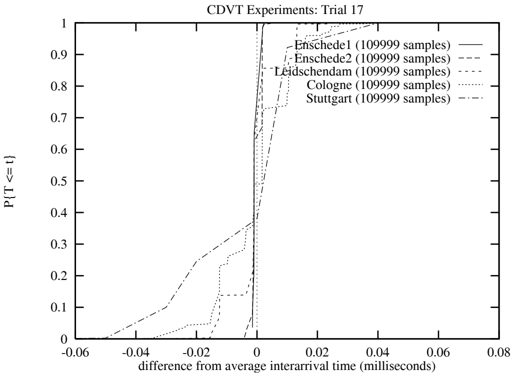


Figure: Empirical CDF: Trial 17: 500 kb/s

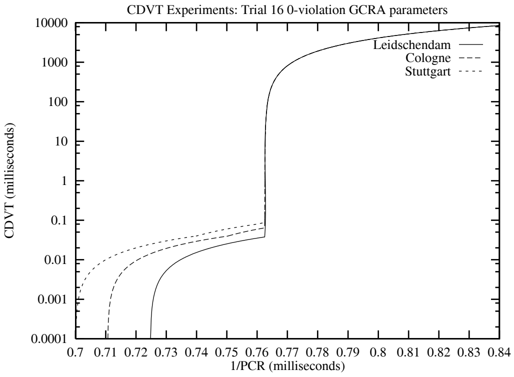


Figure: GCRA Analysis: Trial 16: 500 kb/s

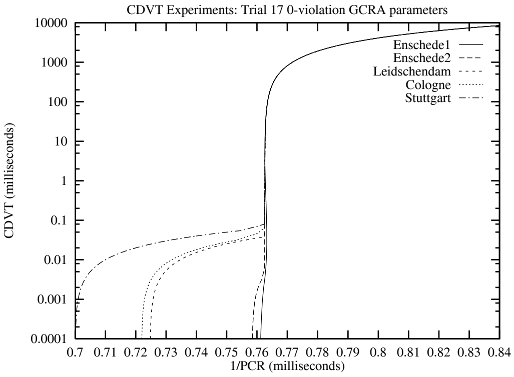


Figure: GCRA Analysis: Trial 17: 500 kb/s

5.8 Assessment of ATM/VBR class of service

5.8.1 Experiment Leader

Olivier Martin, CERN, CH

5.8.2 General explanations about CBR and VBR classes of service

ATM Forum User-Network Interface Specification 3.1 describes the Generic Cell Rate Algorithm (GCRA) algorithm and its relation with the continuous-state leaky bucket (Annex 1 of ITU recommendation I.371). The execution of the GCRA algorithm is governed by two parameters, the increment I (cell inter-arrival time) and the limit L (maximum acceptable deviation).

GCRA(I,T) works as follows:

For every new cell the arrival time is compared to the Theoretical Arrival Time (TAT),

1. if the cell is late (i.e. the cell rate is lower than its nominal rate) it is accepted by the ATM network
2. if the cell is early a check is made to make sure that the deviation is within the tolerable limit, if so the cell is accepted by the ATM network otherwise the cell is discarded.

Although the GCRA algorithm looks very simple and straighforward it is referred to in different ways by the ATM Forum and the ITU, both the increment and the limit can be expressed in different units and using different notations, thus, in the end, generating a great deal of confusion.

The GCRA is used to define, in an operational manner, the relationship between Peak Cell Rate (PCR) and Cell Delay Variation Tolerance (CDVT) and the relationship between Sustained Cell Rate (SCR) and Burst Tolerance (BT).

The CDV tolerance is defined in relation to the Peak Cell Rate but is not a traffic descriptor (i.e. cannot be specified by the user). There is actually an assumption that PCR sources strictly conform to the PCR rate (i.e. GCRA(I,0)) and that CDVs are added by local and/or public ATM switches.

This is indeed the source of one of the major difficulties with public ATM networks that parameters like CDVT cannot be negociated and are instead fixed by the network operators in a way which is not necessarily compatible with default values suggested by some common local ATM switch providers. Although the ATM Forum recognizes that a common, maximum cell delay variation for private, public and hybrid networks is essential (Appendix A - Quality of Service Guidelines), they can only suggest that the receiver CDV tolerance should be designed to handle the case where a connection traverses three networks, each having three switches in tandem. Bellcore TA-NWT-001110 issue 1 proposes an objective value of 750 microsecond for absorbing the cumulative CDV for both DS1 and DS3 circuit emulation services. In Europe, commonly used CDVT values on the European ATM pilot and JAMES (?) are 106 cells (283 microsecond at 150 Mb/s) or 999 cells (2831 microsecond).

It is rather intuitive that as the CDVT increases (i.e. as the ATM network is more tolerant with respect to strict spacing of cells) the traffic can temporarily be more bursty, that is more cells can be transmitted at the peak rate (i.e accepted into the leaky bucket). Therefore, there is a formula to compute the maximum number of contiguous cells (i.e. bursts) that can be transmitted at the full 150 Mb/s link rate while still being compliant with the traffic profile.

For example, to support bust of 30 contiguous cells (i.e. 1440 bytes of payload) at a PCR of 6 Mb/s (mean inter-cell arrival time of 70.66 microseconds) a CDVT of 696 microseconds must be supported by the network.

The VBR service adds another level of flexibility by allowing the user to describe the cell flow of an ATM connection in greater detail than just the PCR aand by distinguishing between Real-Time VBR (rt-VBR) and non-real time VBR (nrt-VBR). In our tests we were only concerned by nrt-VBR which we refer to as VBR.

The aim of VBR is to benefit the service provider as well as the user (i.e. reduced price for equivalent cell rates).

A VBR service is characterized by the Sustained Cell Rate (SCR), Burst Tolerance (BT) and the Peak Cell Rate (PCR), the CDVT associated with the PCR is assumed to be zero. The Burst Tolerance is conveyed in the signalling message in the form of the Maximum Burst Size (MBS) expressed in cells and is derived from the MBS by the formula BT=(MBS-1)*(Ts - T).

For example, in order to support an MBS of 192 cells (9216 bytes) over a VBR service having 6 Mb/s peaks (64 microsecond 48 bytes cell spacing) and 2 Mb/s average (192 microsecond 48 bytes cell spacing), the resulting BT needs to be:

191*(192-64)=24,448 cells

Many different cell flows can conform with the same VBR profile, but the main thing to kep in mind is that on average a VBR flow cannot exceed the SCR, so a good way to visualize a VBR flow is to think about a periodic cell stream with period B*T transmitting B cells at the peak rate and an inter burst spacing at B*(Ts - T) + T, e.g. 2 cells at 6 Mb/s every 320 microseconds or 192 cells at 6 Mb/s every 24,640 microseconds, instead of 1 cell every 192 microseconds (mean SCR).

In practice, things are further complicated by the fact that the user can specify in the signaling message the Cell Loss Priority (CLP), 0 or 1, associated with a given PCR and SCR. The network will drop CLP1 cells in preference to CLP0 cells. It is, in principle, possible to ask for a service such as PCR0=1Mb/s, PCR0+1=2Mb/s, SCR0, SCR0+1.

Appendix B of the ATM Forum User-Network Interface Specification 3.1 provides very interesting conformance examples of traffic contracts for SMDS, Frame Relay and LAN interconnection style of applications. For both the Frame Relay Service (FRS) and the LAN interconnection case, it is suggested to make use of CLP1 cells in order to support bursty traffic and to request the network operator to tag rather than drop non-conforming cells.

The purpose of the study was to determine whether it was possible to make effective use of a VBR service between IP routers interconnected through a public ATM network (JAMES) and how such a VBR service compared with a CBR service having equivalent parameters. Unfortunately it has not been possible so far to organize a truly international VBR connection across the JAMES network. Although many countries do have plans to introduce VBR on their national ATM networks, either pilot or commercial, very few have done so to date, therefore even though the JAMES network does provide VBR it is nearly impossible to access the service!. The only exception appears to be between The Netherlands and Germany and we do plan to organize a VBR test soon.

Two series of tests were made, the intial test were conducted in the Netherlands proved that a Cisco router could be configured in VBR mode and that the resulting throughput was compatible with the definition of VBR, namely that the throughput can never exceed the SCR.

A similar test was conducted in Switzerland during the first 2 weeks of March 1997 in the context of a Technology eXchange Program (TxP), itself a follow-up program of a 155 Mb/s metropolitan ATM project named GENEVA-MAN, where Swiss Telecom actively cooperates with selected users such as the ITU, the University of Geneva, and CERN,

The results of the second test are compatible with those of the first test in the sense that they confirm that Cisco ATM Interface Processors are able to shape IP traffic originating from a LAN in a VBR manner and that there are NO obvious advantages in configuring the interface in VBR rather than PCR mode.

On the contrary, the tests showed that if the network operator accepts to tag traffic beyond SCR with CLP1 instead of dropping cells, it can be very advantageous for the user to configure its Cisco access router in CBR mode with a PCR equal to the PCR0+1 of the VBR service offered by the network operator.

Another way of expressing this is that Cisco does not provide options allowing to make use of a VBR0+1 service with cell tagging.

What is really needed by the academic community is Frame Relay like services where it is possible to burst at access port speeds when there is unused capacity inside the network. However, it is worth noting that the ATM Forum proposed mapping to support Frame Relay service over ATM does not completely replace the functionality of Frame Relay in terms of forward and backward congestion error notifications (FECN, BECN).

In any case, more studies and experimentations are required on real international ATM networks with well documented classes of services in terms of the parameters affecting quality of service and prices.

5.8.3 Detailed Test Results

Environment

CERN: hpstats (traffic generator), FDDI connected (CIXP) to:

Cern-atm7: Cisco 7010

```
IOS (tm) GS Software (GS7-J-M), Version 11.1(4)
AIP 2, hardware version 1.2, microcode version 10.13
FIP 0, hardware version 2.9, microcode version 10.2
```

Fore Switch: ASX200-BX

```
Hardware version 1.0, Software version S_ForeThought_4.0.2 (1.15)
2*NM-C-OC3c/STM1c-TIMING-MM-SC-128KB-4PT (Rev. 1.1)
1*NM-C-E3-TIMING-128KB-2PT (Rev. 1.0)
```

Traffic parameters: conf port traffic

Port configuration:

```
CBR  VBR  ABR-UBR  Port  Qsize  CDV  Qsize  CDV  QsizeEFCI-ON EFCI-OFF
1A1  256  700      256   1400   256  64     1
1D4  256  700      256   1400   256  64     1
```

Port priority queues:

```
CLP   Qsize     Qsize
Port  Priority  Threshold  Dedicated  Current  TxCells    LostCells
1A1   VBR       256        256        0        9533674    0
1D4   VBR       256        256        0        22722228   0
```

upc parameters: conf upc

```
Index  PCR01  SCR01  MBS01   PCR0   SCR0   MBS0   CDVT Act  EPD Name
0                                                      drop no  default_ubr
1      15625                        5208     30        drop no  SwissPTT
3      15625                        5208     32        drop no  vbr0
4      15625                        5208    255        tag  no  vbr0
```

atm vp parameters: (conf vp)

```
Input       Output
 Port  VPI  Port  VPI   MaxBW      BW MaxVCs  VCs  UPC  Prot
 1A1     1  1D4     2    6.6M   0.0K    N/A  N/A    4  pvc
 1D4     2  1A1     1    6.6M   0.0K    N/A  N/A    4  pvc
```

Swiss Telecom: rtr-ccatm

```
cisco 4700 (R4K) processor (revision E) with 16384K/4096K bytes of
memory
IOS (tm) 4500 Software (C4500-I-M), Version 11.1(6)
ATM Unit 0, Slot 1, Type ATMizer BX-50, Hardware Version 1 ATM Xilinx
Code, Version 1, ATMizer Firmware, Version 2.0
```

Tests have been conducted with netperf between hpstats (CERN) and a SUNOS IPX station (Swiss Telecom), most tests have been made in the direction CERN--->Swiss Telecom;

Fore Switch: ASX1000 with 2*switch fabric (ASX200-BX equivalent) Software version S_ForeThought_4.0.1 (1.20)

Sparc/4 station with SunOs 5.5.1

CDVT 700 microsecond (250 cells), leaky bucket buffer size 256

First round of tests was conducted with upc 1, PCR01 15625 (6Mb/s), SCR0 5208 (2Mb/s) and MBS0 30 on the 3 Fore switches while the Cisco were configured with atm pvc 150 1 150 aal5snap 6000 2000 1 (i.e. peak rate 6Mb/s, sustained rate 2Mb/s, maximum burst 32 cells).

Not surprisingly the non-matching bursts do cause cell losses and therefore packet losses, teherfore like for CDVT user and network configured MBS parameter values must be compatible (I know it is obvious but there is no harm in repeating the obvious).

The MBS were adjusted the day after to 192 cells (6*32 on the Cisco) and the following results were obtained:

Cisco atm pvc command:

Peak	Sustained	Burst	Throughput (Mb/s)
6000	2000	6	0.05
6000	2000	5	0.63
6000	2000	4	1.47
6000	2000	3	1.47
6000	2000	2	1.55
6000	2000	1	1.50
2000	2000	(CBR)	1.53
2200	2200	(CBR)	1.61/1.68
2500	2500	(CBR)	0.67
3000	3000	"	0.23

So there seems to be a problem between Cisco and Fore when the MBS values are too close. The problem was traced later to be a Cisco 4700 problem only.

We also noticed that traffic between CERN and Swiss Telecom was slightly but repeatedly faster than traffic from Swiss Telecom to CERN. The tests were repeated the day after with an MBS of 1920 cells (i.e. 60*32 cells)

Peak	Sustained	Burst	Throughput (Mb/s)
6000	2000	60	1.64
6000	2000	61	0.24
6000	2000	58	1.58/1.64
6000	6000	(CBR)	4.44

The combination 6000/2000/60 at CERN with 6000/2000/58 (Swiss Telecom) works (Cisco 4700 problem).

The tests were repeated the day after with an MBS of 255 cells (as it is now well established that the burst size has essentially NO effect on the average throughput which always conforms to the sustained cell rate we did not bother any more about the MBS parameter on the Cisco side).

Peak	Sustained	Burst	Throughput (Mb/s)
6000	2000	7	0.11
6000	2000	6	1.58
6000	2200	6	0.11

6000	2100	6	0.11
2200	2200	(CBR)	1.75 drop
2400	2400	(CBR)	0.64 drop
3000	3000	(CBR)	0.24 drop
2400	2400	(CBR)	1.91 tag
3000	3000	(CBR)	2.38 tag
4000	4000	(CBR)	3.16 tag
5000	5000	(CBR)	3.96 tag
6000	6000	(CBR)	4.71 tag
6600	6600	(CBR)	5.17 tag
7000	7000	(CBR)	0.15 tag

N.B. Tagging not enabled on the Swiss Telecom side (only on the sending side) However, we got similar results with tagging also enabled on the Swiss Telecom Fore switches.

3000	**3000**	**(CBR)**	**2.38 tag**
4000	4000	(CBR)	3.17 tag
5000	5000	(CBR)	3.96 tag
6000	6000	(CBR)	4.70 tag
6600	6600	(CBR)	5.16 tag
7000	7000	(CBR)	0.15 tag

The above tests confirm the results already obtained by Surfnet (see D11.2) as well as the definition of the VBR class of service, namely that it is never possible to achieve throughputs higher than the SCR, and therefore that there is NO interest to shape non-VBR traffic sources, such as Cisco routers aggregating IP traffic from local/regional/national networks, in a VBR manner in order to access an ATM network, unless the VBR service is priced more attractively than the equivalent CBR service.

The only way to achieve higher throughput than the SCR of the VBR service is to configure the Cisco router in CBR mode and request the network operator to tag cells between SCR and PCR with priority 1. In other words, configuring CBR on the user side in order to access a VBR service with the tagging option enabled seems to be the best choice. But, we have no knowledge of whether such services will be available from public ATM operators and if so at which cost.

On can also hope that if such services were available, ATM access providers such as Cisco would allow to make use of the service in a more natural manner than using CBR over VBR. The tests obviously need to be repeated with multiple rather than single traffic sources in order to understand better the behaviour of the public ATM network and TCP.

Finally we should make more testing with the Early Packet Discard (EPD) option enabled. This is planned for the second phase of the experiements.

5.9 Performance of the Native ATM Protocol

Due to a lack of native ATM applications, we were not able to test this. The experiment was deferred to phase two.

5.10 IP resource reservation over ATM

5.10.1 Experiment Leaders

Sabine Kuehn, University of Dresden, Germany and Olav Kvittem, UNINETT, Norway

5.10.2 Summary of results

University of Dresden has developed RSVP over IP over ATM for DEC Workstations so that this functionality could be tested in a local ATMenvironment using an own performance tool with an integrated graphicalRSVP user interface. Moreover, the U of Dresden is developing a videoconferencing system testing RSVP over ATM by a more practical-relevantexample.

An introduction talk to RSVP was held at a TF-TEN meeting in Stuttgart 29/1.This is available via the TF-TEN homepage. RSVP/IP implementations has been investigated and sucessfullyinstalled.

5.10.3 Further participants

Frank Breiter, Technical University of Dresden, Germany

5.10.4 Dates and phases

The project was prolonged for 4 months in order to get broader experiences in operational requirements in a pilot experiment and to find resources to completethe project.

Revised plan	**Dates**	**Results**
1. Investigation	96-07 - 96-10	
2. Initial experiments	96-10 - 96-12	detailed pilot documentation
3. Pilot experiment	97-01 - 97-03	operational infrastructure
4. Reporting	97-03 - 97-04	report

The pilot experiment has not been completed due to delays by a subcontractor from UNINETT.

5.10.5 Network infrastructure

The initial tests were done in a local ATM environment at thetechnical university of Dresden.

Our experimental environment consists of several multimediaworkstations of type DECstation 3000 AXP 700 and 300 which areconnected with a DEC Gigaswitch/ATM via multimode fibre. The celltransmission is performed using SONET/SDH frames with standard 155Mbps per channel. Only AAL5 is currently implemented. The switch andATM adapter cards support UNI 3.0 signalling as well as UNI 3.1 andoffer NSAP/E.164 addressing. Moreover, CBR (constant bit rate), andABR (available bit rate), both with point-to-point andpoint-to-multipoint VCs, are possible. Our local environment iscurrently neither connected to JAMES nor to any other public (B-WiN/DFN) or private ATM networks.

Waiting for SVC-project to provide basics for a SVC-infrastructureover the overlay network for performing the pilot experiment.

5.10.6 Results and findings

1. Investigation phase

Engineering issues includes how to realise RSVP over IP over ATM as the different concepts make an integration of RSVP and ATM even more difficult. There are some outstanding issues like: how to make dynamic QoS changes for existing VC (maybe without establishing of a newVC?).

Certainly there will be more than one approach in realising RSVP over IP-ATM. To avoid considerable changes in RSVP we propose the following way: receiving a reservation message from the downstream host, the appropriate router or host establishes an ATM connection to the downstream hop according to the reservation information. On this basis, it will also be possible to establish ATM point-to- multipoint connections, at least to homogeneous receivers. ATM presumes homogeneous receivers even in case of heterogeneous RSVP-reservations, therefore routers have to reserve according to the highest reservation requirements. Reserving VC's between routers in an ATM network depends on classical IP over ATM model (ARP) in case of more than one LIS. However, a realised NHRP over ATM would allow to establish ATMshortcuts without any changes of RSVP. So an extensive modification of RSVP to realise ATM shortcuts in combination with ARP will be unnecessary. The considerable differences between the service classes of the Integrated Services IP and ATM also require detailed analysis of mapping service classes as well as traffic and quality of service parameters. Translating such kinds of parameters is an additional service for the layer-to-layer communication during the call establishment phase.

2. Practical aspects

There are RSVP implementations available for IP on various UNIX'es from ISI and from release 11.2 on Cisco routers. At the U of DD there exists a modified ISI-RSVP implementation for DEC-host and router (Digital Unix), with an integrated functionality of mapping RSVP on ATM VC's.

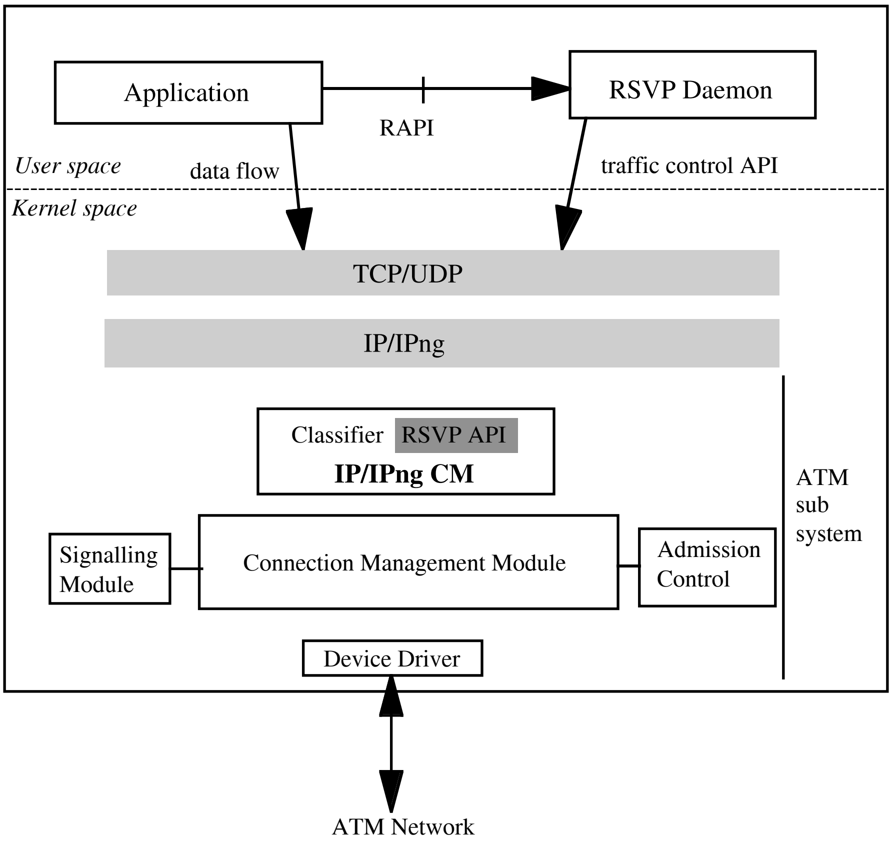


Figure 1: Interaction between ATM and RSVP (DU)

It uses the ATM subsystem as a part of the operating system and offers a specific API as a part of the convergence modules. With IP over ATM, this API is available in the IP convergence module and supports realtime handling and transport of IP flows (figure 1). The API is alsoable to receive reservation information from a local daemon as a partof the RSVP implementation. Based on information contained in a given flow specification, a new ATM VC reservation is performed after completing the mapping of service classes and parameters. Moreover, after calculating and adding the appropriate overhead (e.g. of the AAL5 trailer) to traffic parameters, a conversion of RSVP parameters (e.g. [bit/s]) into ATM parameters (e.g. [cells/s]) is necessary. Aclassification of IP flows belonging to a dedicated reserved VC is done in the convergence module. This is realised based on the classification of the IP source address/port pair, contained in the IP packet, to the corresponding reservation (virtual channel). This implies the fact, that packets which cannot be identified based on filter specification information will be transmitted over an available bit rate channel only.

3. Initial experiment

Only unicast tests were performed in 1 LIS over ATM which have resulted in the establishment of a reserved pt-2-pt VC between sender and receiver situated in the ATM net. Testing the behaviour of a RSVP router connecting two LIS which were set up over the ATM net, two single VC (between receiver and router; router and sender) were been established, able to carry data. The used application on top of RSVP were: rtap and a modified performance tool with an integrated graphical RSVP interface (surface).

4. Suggestions for a pilot project

a.) Test of a RSVP based video tool e.g. VIC or NV as part of Mbone. Transmission of RSVP messages over ATM SVCs using ARP to get e.g. information about interoperability of exchanging RSVP messages between several RSVP implementations (CISCO and ISI)

b.) The same tests as in the local environment should be repeated over JAMES with SVC tunnelling

c.) Eventually, if there is enough time to implement a Multicast Address Resolution Server multicast experimentscould be performed.

In preparation for the pilot experiments measurements/investigations should be stated e.g. the increase of the establishment time caused by the additional exchange of RSVP messages, the behaviour of the RSVP protocol and the interaction of RSVP and ATM over a large net. The tests were planned between Norway and Germany using DEC equipment as we do not know about any other RSVP over ATM implementations.

5.10.7 Test related problems and general comments

There were some initial problems getting ATM SVCs to work.

University of Dresden has a public ATM connection (DFN: B-WiN) not the Faculty of Computer Science itself, and it turned out that they could not get a JAMES admission with their DEC equipment?

In Norway the intended DEC equipment was not made available as expected. Further studies for the mapping RSVP multicast to ATM are done and to add/drop party. In view of a practical realisation of a RSVP based video-conferencing between multiple receivers we are currently working on a MARS for Digital Unix.

5.11 Security in ATM Networks

5.11.1 Experiment leaders

Paulo Neves and Roberto Canada

5.11.2 Goals

We intend to present a report on:

- current ATM specifications;
- how JAMES network is vulnerable or reliable, and;
- what can be done to improve security over JAMES.

To achieve these goals it is necessary to:

- determine in which way does JAMES comply with current ATM specifications on security;
- test vulnerability of ATM networks to several kinds of attacks and;
- point out security services that ATM networks should/could provide

5.11.3 Phases of the experiment

Phase 1: Definition of the experimental framework

- General network security requirements;
- Possible threats;
- Basic Security Services in ATM networks;
- Currently available specifications, regarding security;
- What is JAMES providing?

Phase 2: Planning of experiments

- What tests can be done over JAMES;
- Reliability and Fault Tolerance tests;
- Other tests.

Phase 3: Experiments and data collection

5.11.4 Network infrastructure

For most of the experiments, a UNI compliant ATM interface would be required, with accessible Control and Management Planes of the Protocol Reference Model [8] on the intermediate switches. As that is not yet available in the JAMES context, the use of SVCs over PVCs [1] is being considered.

5.11.5 Results and findings

The initial phase of the workpackage has established the following points:

1.Security requirements of communication networks include:

- Availability
- Secure communication channel
- Accurate auditing information

We consider that aspects like user authentication and non-repudiation of contents (of user messages) should not be expected from the network as an entity, although they might be supported by other means.

2. Threats Analysis

Three classical attacks and their consequences on each ATM flow were studied, to deduce flow's vulnerability:

- data or traffic flow confidentiality loss due to an intruder eavesdropping the network and deducing user data content or user traffic features
- data integrity loss caused by accidental or malicious injection/removal/modification of cells/signalling messages in transfer
- overloading problems following a mass-injection of cells/signalling messages.

Overloading consists in disrupting network entities (e.g. ATM switch) or end-entities (e.g. end-station) by sending a large number of cells/signalling messages whose processing prevents other useful cells/messages processing or at least slows it down. This attack is particularly serious when done with SET UP messages and is also known as Denial of Service (DoS).

2.1 User data flow

Confidentiality and integrity losses are particularly damaging when applied to user data flows since an intruder eavesdropping at a point on the network can retrieve all the cells belonging to one connection (i.e. carrying the same VPI/VCI values), evaluate the amount of information transmitted and even deduce their content after having assembled cells back. Eavesdropping appears as a serious problem especially when applied to sensitive data transfer.

An intruder may also disrupt the network by injecting, modifying or removing user cells. Most often these cells are removed at the receiving entity (because they fail the upper layers integrity check), causing retransmission of upper-layer frames, and overloading the network. In other cases, some of them may be processed and disastrous consequences may happen (when, for instance, a financial transaction transfer is performed).

2.2 Signalling flows

Signalling flows ([10],[9]) vulnerability is message type dependent. Since SET UP messages for establishing point-to-point connection are the only ones bearing the sensitive information - called and calling end-entities addresses, they appear as the most vulnerable messages to eavesdropping attacks. Indeed an intruder wanting to identify the communicating entities has only to eavesdrop the signalling flow during connection set up. Retrieving their identities can be of interest for him, but additionally he can capture the returned CONNECT or CALL PROCEEDING message which includes the VPI/VCI identifiers assigned by the network to the new connection and then eavesdrop the corresponding user channel (VPI/VCI) to infer exchanged user data.

Also overloading the network with SET UP messages is damaging since this causes mass connections set ups and therefore end or network entities overload and consequently legitimate connections rejections.

Other messages such as RELEASE and RELEASE COMPLETE are vulnerable to integrity attacks because their injection immediately causes a connection release, which can also be viewed as a DoS attack.

2.3 Management flows

Management flows ([10],[11]) are especially vulnerable to confidentiality and integrity attacks. An intruder eavesdropping performance management cells can infer the number of user cells transmitted over one connection. Also an intruder realising an attack on integrity may cause line errors to remain undetected (by removing AIS/FERF cells or injecting continuity check cells), a connection release whereas the connection is still operational (by injecting AIS/FERF cell, removing continuity check cells or modifying performance management cells with a significant increase of the transmitted errored cells number or the total number of transmitted user cells) or a bad line problem location (by tampering AIS/FERF cells).

3. Security services requirements for ATM

Considering the results of the preceding points, summarised in Table 1, security services need to be introduced within ATM planes to protect ATM flows exchanges (see Table 2).

	user data flows	**signalling**	**management flows**
data and traffic flow confidentiality	disclosure of data (exchanged over one VPI/VCI connection)	disclosure of the communicating parties identities and VPI/VCI associated to the connection	disclosure of the amount of user data exchanged
integrity	tampered cells processing	connection release	connection release
overloading	useful cells processing prevent	multiple connection set ups	useful cells processing prevent

Table 1

user plane	**signalling plane**	**management plane**
confidentiality	---	confidentiality
integrity	integrity	integrity
replay detection	replay detection	replay detection
padding (against traffic flow confidentiality attacks)	---	---

Table 2

3.1 Signalling plane

Protecting signalling flows against integrity and overloading attacks requires the introduction of authentication, integrity and replay detection services, naturally complemented by access control mechanisms. Note that not only end-entities (end-stations) but also network entities (switches) need to handle these security services for detecting bogus RELEASE or SET UP messages.

3.2 User plane

User data flows are vulnerable to data confidentiality, traffic flow confidentiality and integrity/overloading attacks so that respectively confidentiality, padding and authentication/integrity/replay detection services must be introduced within user plane.

3.3 Management plane

As shown in table 2, management flows need the introduction of confidentiality, integrity, access control and replay detection services. Note that, in case management cells' content is encrypted, the integrity service is naturally performed thanks to the management cells' CRC field ([10],[11]) being encrypted along with management information. On the other hand, given the fixed management cells structure with only 6 bits being free (the "reserved" field), replay detection seems impossible to realise.

4. Availability

We consider the availability of some of these services (namely to the Control and Management Planes) is essential for the robustness of the network itself. In fact, we find that the integrity of the network depends on the existence of means to avoid some forms of attack (Denial of Service, Masquerade, Spoofing and Repudiation), on signalling and management protocols, even if user security services could be performed at higher layers.

5. Standardisation

Standardisation work at the ATM Forum is under way regarding the future shape of ATM Security infrastructure [5]. This infrastructure considers the use of special signalling procedures to allow for negotiation of security parameters between communicating parties.

6. JAMES framework

In the JAMES framework we are confined to user data channels, running through PVCs, without any means to directly contact intermediate ATM switches, for connection negotiation or management. In order for us to test the most interesting issues, some control and management functions would have to be present.

5.11.6 Relevance for service and migration suggestions

Given the above considerations, we think security tests in which robustness of the network to attacks is verified is pertinent in view of a future ATM production network in Europe. The results gathered would be useful in establishing what are the exact requirements for security in such an environment, and allow a comparison between these and the ones already proposed by the standardisation bodies.

5.11.7 Further studies

As soon as a true UNI for JAMES is in place, we can proceed with our field tests, simulating the following attacks:

- Masquerade;
- Protocol spoofing;
- Denial of Service;
- Repudiation.

In the meantime (as soon as our connection to JAMES is established) we will try to develop some experimental work over the SVC infrastructure.

5.11.8 Acknowledgements

We had a valuable contribution from Maryline Laurent and Pierre Rolin [12] on which this work is partly based on.

5.11.9 References

[1] Graf, C., "SVC Tunnelling through PVPCs - V. 3", TF-TEN, 8 August 1996

[2] Chuang, S-C., "Securing ATM Networks", 18 October 1995 (Interim report)

[3] Peyravian, M., "ATM Security Scope and Requirements, ATM Forum/95-0579

[4] Laurent, M., Rolin, P., "Etat de l'art de la securiti sur ATM", DNAC'96

[5] Tarman, T. D., "Phase I ATM Security Specification", ATM Forum/95-1473R2, 15 April 1996

[6] Peyravian, M., "A Certification Infrastructure for ATM", ATM Forum/95-0964

[7] Peyravian, M., "A Framework for Authenticated Key Distribution in ATM Networks", ATM Forum/95-0580

[8] ITU-T, "I.321 - B-ISDN Protocol Reference Model and its Applications", Geneva, April 1991

[9] ITU-T, "I.311 - Integrated Services Digital Network (ISDN): Overall network aspects and functions. B-ISDN general network aspects", March 1993.

[10] ATM Forum, "ATM User-Network Interface Specification", version 3.1, 1994.

[11] ITU-T, "I.610 - Integrated Services Digital Network (ISDN): maintenance principles. B-ISDN operation and maintenance principles and functions", March 1993.

[12] Laurent, M., Rolin, P., "Securite ATM: une analyse de flux menee sur quatre architectures de reseaux", GRES'95, Paris, September 1995;ftp://ftp.rennes.enst-bretagne.fr/pub/security/ml_GRES95.ps.gz.

Glossary

ARP	Address Resolution Protocol
ATM	Asynchronous Transfer Mode
CBR	Continuous BitRate (ATM Forum: traffic class)
DCC	Data Country Code
DBR	Deterministic BitRate (ITU-T: traffic class, eq CBR)
E.164	(ITU-T addressing standard)
ICD	International Code Designator
IESG	The Internet Engineering Steering Group. Manages the working groups and standardization process in IETF
IETF	Internet Engineering Task Force (http://www.ietf.org) The Internet protocol standardization body
ILMI	Interim Link Management Interface
IP	Internet Protocol
ISO	International Standards Organisation
ITU	International Telecommunications Union
JAMES	A European experimental ATM-network.
LIS	Logical IP Subnetwork
MBS	Maximum Burst Size (ATM Forum: traffic parameter)
NHRP	Next Hop Resolution Protocol (ftp://ds.internic.net/internet-drafts/draft-ietf-rolc-nhrp-11.txt)
NHRP-R2R	NHRP for Destinations off the NBMA Subnetwork ftp://ietf.org/internet-drafts/draft-ietf-ion-r2r-nhrp-00.txt
NRN	National Research Network
NSAP	Network Service Access Point (OSI term)
OAM	Operations And Maintenance
PCR	Peak Cell Rate (ATM Forum: traffic parameter)
P-NNI	Private Network to Network Interface
PNO	Public Network Operator
PVC	Permanent Virtual Circuit
PVPC	Permanent Virtual Path Connection
RSVP	Resource ReSerVation Protocol Version 1 Functional Specification - Internet draft; http://www.internic.net/internet-drafts/draft-ietf-rsvp-spec-12.txt
SBR	Statistical BitRate (ITU-T: traffic class, eq VBR)
SCR	Sustainable BitRate (ATM Forum: traffic parameter)
SNMP	Simple Network Management Protocol
SVC	Switched Virtual Circuit
TCP	Transport Control Protocol
UDP	User Datagram Protocol
UNI	User Network Interface
VBR	Variable BitRate (ATM Forum: traffic class)
VC	Virtual Circuit
VP	Virtual Path
VPC	Virtual Path Connection