Project Number: RE 1009
Project Title: TEN-34

Deliverable D11.2

Interim Results of Phase 1 Test Programme

Deliverable Type: PU - Public
Contractual Date: 31 December 1996
Actual Date: 20 December 1996
Work Package: 11 - Phase 1 Test Programme
Nature of Deliverable: RE - Report

Authors:
Michael Behringer, DANTE
Mauro Campanella, INFN/CNAF
Zlatica Cekro, ULB/STC
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. This deliverable describes the interim 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 interim results of the experiments have shown that basic ATM services such as CBR are well understood and can be used in a production environment. In the area of the more advanced ATM services the interim results demonstrate that the services vary from being usable only under very idealistic circumstances (e.g., switching), to potentially providing a stable infrastructure, pending further investigation (e.g., VBR).

The most important result of the experiments so far is that switching proves to be very difficult to use in the way it was created for, which is between end-user applications. There are no major problems with switching in a LAN environment, where bandwidth is close to unlimited. But over a public wide area network, with significant bottlenecks and policing, SVCs can only be used under severe limitations. Due to the number of diverse problems it is not expected to have SVCs available in a general production environment within the next two to three years.

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 fast 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 need to be carried out for the development of future services. This deliverable reports the first interim results of these experiments. A brief summary of the interim results per experiment can be found below.

The experiments carried out so far show that conservative ATM traffic classes such as CBR can be used without major problems to provide a transmission infrastructure for IP based traffic. Essentially the characteristics of an ATM CBR service are similar to that of a leased line, if the ATM overhead is not taken into account. It is believed that also VBR services support IP based platforms without major problems, but this could not be verified yet, as JAMES did not offer a VBR service as of the time of writing this report.

The next challenge for ATM lies in the more complicated switching of virtual circuits (SVC), and on services which require or make use of switching, such as resource allocation with RSVP. In this area we discovered several problems, which are mostly due to missing implementations and standards. Whilst in the local area network bandwidth is close to unlimited, in the wide area network there are likely to be narrow bottlenecks, which are policed against over-usage. One of the main problems is that there is currently no means of negotiating bandwidth between end-user applications and the underlying network. Also it is not clear if switches will be able to handle a high number of SVCs such that the traffic shaping does not violate the policing functions of the public network. There are possibilities to circumvent some of these problems by hard-coding certain parameters, which would normally be handled transparently. But for the general case it is unlikely that switching will be usable for a few more years.

The problems discovered with switching also had impact on some of the other experiments depending on SVCs. Therefore the experiments on ATM ARP, RSVP and NHRP have been delayed. We still expect to get results for these experiments within phase 1 of the testing programme.

More advanced ATM traffic classes such as ABR and UBR over a wide area network have not been considered in the first phase, because they are mostly not available yet. These services will be examined in phase 2 of the test program (work package 14).

The JAMES network offers up to today only very basic ATM services such as CBR, which are far too conservative for testing advanced ATM services. This is a serious restriction for some of the experiments planned here. No service level guarantees such as set-up times for VPs are available, so that it is very difficult to make detailed project plans. To some extent these problems could be circumvented by requesting a large static set of VPs, which can be used as required by TEN-34. More advanced ATM services such as switching are carried out on NRN equipment, using the JAMES network only as transmission infrastructure. The lack of VBR cannot be solved by other means, but we were assured that VBR services will become available in some countries from December 1996.

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 usable bandwidth.

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 transparently that is not capable of switching. The second result is that there are a number of technological shortcomings in different areas such that switching between end-user applications in a production environment cannot be envisaged over a wide area network for some years to come. Further investigation is necessary.

3. **Classical IP and ARP over ATM**: Local tests have shown no problems in doing address resolution over a LAN. These tests have yet to be confirmed in the wide area network and with more than one server, but no major problems are expected.

4. **IP routing over ATM with NHRP**: Extensive testing has not yet started because there are no implementations available yet, and because of the problems encountered in the SVC experiments (see 2.), which are required for this experiment.

5. **European ATM Addressing**: Initial investigations showed that most NRNs will use NSAP addressing, while there is no agreement between PNOs on which standard to prefer. The investigation is ongoing.

6. **ATM Network Management**: Experiments with SNMP based network management of NRN routers and switches have shown no problems. Further work is planned to be done on other management platforms such as X.user.

7. **CDV over concatenated ATM networks**: Local tests of the measuring and traffic producing equipment were carried out, which gave a better insight into the technology. The international experiments will start shortly.

8. **Performance of the Native ATM Protocol**: This experiment has not started yet.

9. **Assessment of ATM/VBR class of service**: This test has not started yet, as no VBR service is available over JAMES as of today.

10. **IP resource reservation over ATM**: Local tests were carried out successfully and without major problems. International tests are pending. The delay is mainly due to the delay in the SVC experiment.

11. **ATM Security**: This is a newly defined work item, which was not included in D11.1, the specification of the phase 1 test programme. The initial work focused on a precise description of the security relevant areas in ATM. The work will be focused on the security of the ATM layer itself, mainly related to switching.

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 on offer by JAMES are CBR and SMDS and as of September 1996 an IP service is also offered. Of these only CBR 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). In addition to that more advanced services are needed. For this phase of tests this is mostly VBR and switching. The lack of a VBR class of service cannot be solved by other means by TEN-34, so that the planned VBR experiment will be delayed until a VBR service becomes available. The lack of switching 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 a 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 the VBR traffic class will become available before the end of 1996 in some countries, switching is foreseen to become a service over parts of JAMES during the first half of 1997.

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.

To circumvent these problems we requested a large static set of VPs, so that the bandwidth can be allocated to experiments directly by TEN-34, 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

As there is a significant overlap in the testing programmes of JAMES and TEN-34, some of the experiments, where the interests of the two groups overlap, were identified as joint experiments. The intention was to define in more detail the precise co-operation on a per experiment basis, and to outline a workplan for each experiment between JAMES and TEN-34. This activity took longer than expected, and the joint experiments are currently being defined in detail. The reason for the delay is mainly that it was not easy to define the work distribution between JAMES and TEN-34.

At the last TERENA TF-TEN meeting (the group who carries out the TEN-34 experiments) on 30. and 31. October 1996, a number of technical JAMES representatives were invited, and significant progress was made in getting to understand the mutual interests and possibilities better. Detailed experiment plans for the joint experiments are now being defined between the experiment leaders on both sides.

4. Timing of the Experiments

Most experiments are progressing within the defined time limits. In the case of a few experiments the TEN-34 test programme is running behind the plan set out in the preliminary testing schedule outlined in D11.1 (Specification of Phase 1 Test Programme). In some cases (namely the ARP, NHRP and RSVP tests) this is due to the unforeseen problems with ATM switching, so that subsequent tests were delayed as well. The VBR test could not be started because there is still no VBR service over JAMES.

The testing schedule was originally defined in a very tight time frame, ending in December 1996, while the phase 1 test programme actually lasts until April 1997. This was done to allow for the inevitable delays in testing new and so far un-tested ATM techniques. Especially with the delay in the testing of ATM switching over wide area networks this buffer time proves valuable. We therefore still expect to be able to finish the majority of the experiments planned within phase 1 of the programme. In some cases however the tests require additional resources, the availability of which is not in the hands of TEN-34. This is the case for implementations of NHRP, which are still being developed by router vendors, and for the VBR service over JAMES. In the latter we were re-assured that VBR will be available at least in some JAMES countries by the end of 1996, so that we could meet our target date. In the case of NHRP it is expected that working implementations will be available soon, but there are no guarantees. In the worst case this activity has to be continued in phase 2 of the test programme.

5. Detailed Test Descriptions

The following sections describe the interim results of the experiments in detail. The experiments are:

5.1 TCP-UDP/IP Performance over ATM
5.2 SVC Tunnelling through PVPCs
5.3 Classical IP and ARP over ATM
5.4 IP routing over ATM with NHRP
5.5 European ATM Addressing
5.6 ATM Network Management
5.7 CDV over concatenated ATM networks
5.8 Performance of the Native ATM Protocol
5.9 Assessment of ATM/VBR class of service
5.10 IP resource reservation over ATM
5.11 Security in ATM Networks

Not all of the experiments have concluded. The full results will be published in deliverable D11.3.

5.1 TCP-UDP/IP Performance over ATM

Experiment Leaders

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

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 end-nodes 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 Kbytes) 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 end-nodes.

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.

Participants

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

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.

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: more end-nodes 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.

Network infrastructure

1. 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.

2. Network configurations

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

The load on the end-systems during the experiments were negligible, since dedicated equipment has been used.

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.

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.

The Sun SS5 has a 85 MHz processor featuring 65.3 SPECint and 53.1 SPECfp. It has a FORE SBA-200 ATM adapter with driver version Fore 4.0.

The HP 725/75 has a PA-RISC processor running at 75 MHz. The Fore ATM adapter installed on it runs the driver version 4.0.

The Sun SS20 has a Supersparc CPU running at 75 MHz, 1 MB of external cache, 80 SpecInt92, 104 SpecFp92 and SBus I/O bus at 100 MB/s. Its ATM adapter was the ForeSystem SBA-200 at 155 Mb/s on fibre.

The Silicon Graphics has a R4400sc CPU,, 16 kB + 16 kB internal caches and 1 MB external cache, 140 SPECint92, 131 SPECfp92 and GIO I/O bus at 110 MB/s. Its ATM adapter is a ForeSystem GIA-200 at 155 Mb/s on fibre.

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.

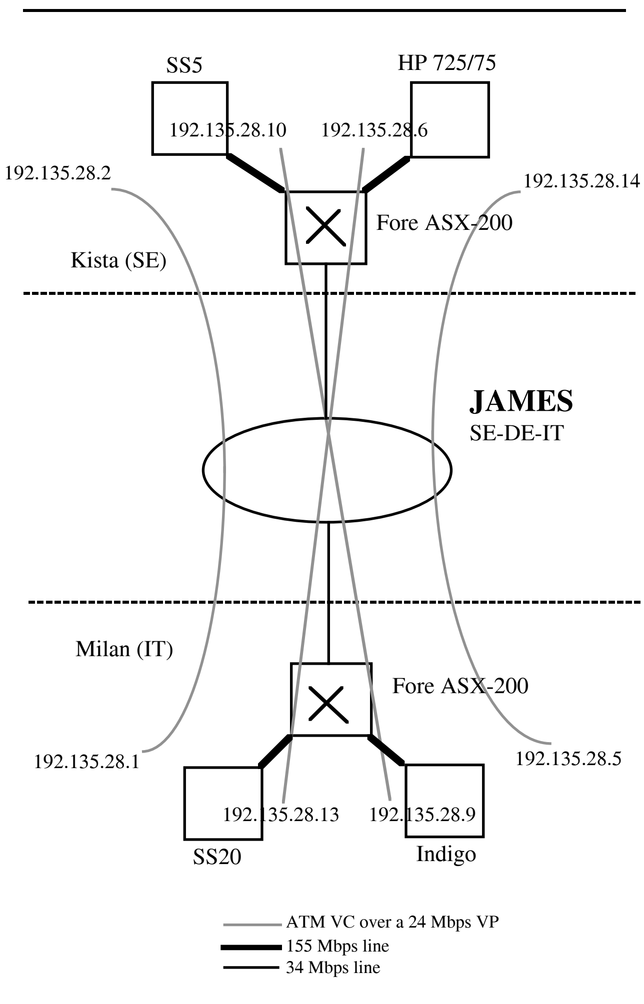

Figure 1: Equipment and network configuration in the test session Italy-Sweden.

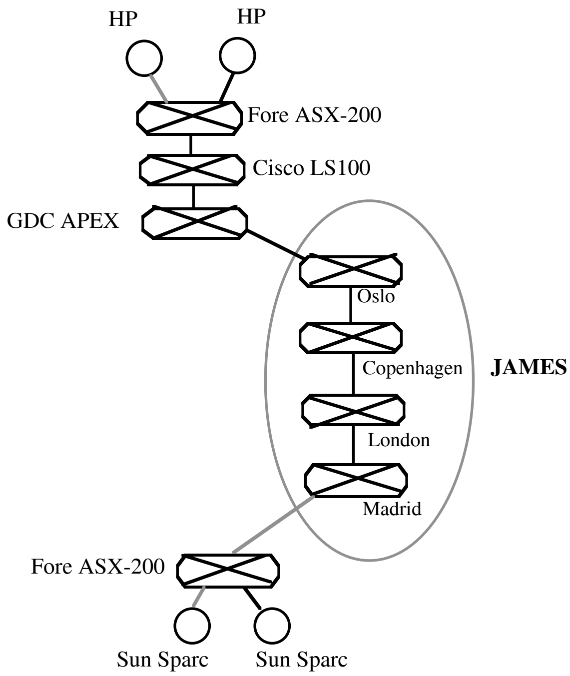

Figure 2: Network configuration in the test session Norway-Spain.

Results and findings

We present in the following the list of some of the most important results 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

1.1 Round Trip Time

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. Ping can be a reliable mean to measure RTT since in our case, the end-nodes involved in the experiment were dedicated workstations without other running user processes.

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.

1.2 Maximum throughput and Window Scaling Option

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 end-nodes. 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 lost, 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-ES):
  Win = (63 msec * 13.824 Mbps) / 8) is approx. 109 kbytes
- or the session Italy-Sweden (IT-SE):
  Win = (48 msec * 24 Mbps) / 8 is approx. 120 kbytes

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 byte the higher bandwidth allocated to the VP, which gives a shorter transmission time.

The maximum window size allowed by traditional operating systems is 64 kbytes, 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 end-nodes 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. In both cases the MTU (Maximum Transfer Unit) used in the experiments was 9180 bytes. 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.

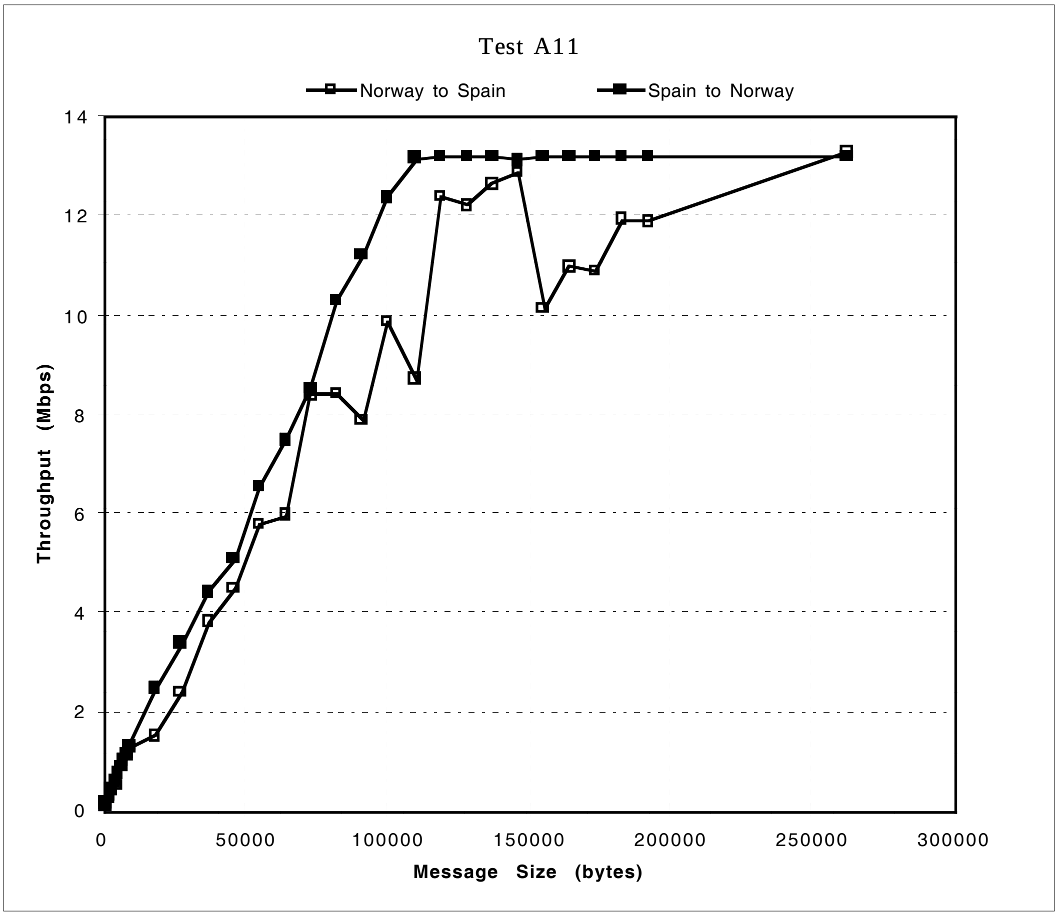

Figure 3: Test session Norway-Spain: Throughput measurement for a one-way TCP connection with variable local ssb/rsb sizes , remote ssb/rsb sizes and message size (ssb = rsb = msg).

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 kbytes, 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.

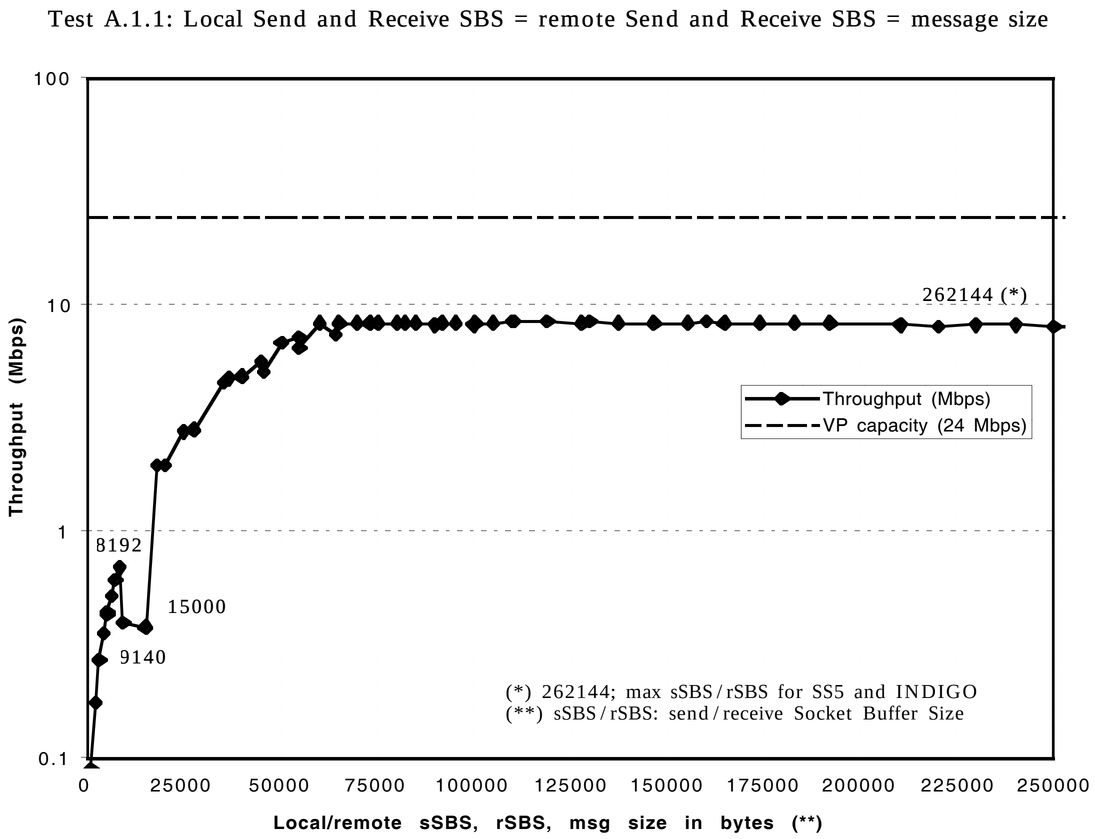

Figure 4: Test session Italy-Sweden: Throughput measurement for a one-way TCP connection with variable local ssb/rsb sizes ,remote ssb/rsb sizes and message size (ssb = rsb = msg).

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 kbytes: 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 kbytes, 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 end-nodes;
- 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.

All the throughput values analysed in the following paragraphs, refer to the application level; this means that all the protocol overheads from the TCP level to the physical level are not taken into account.

2.1 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-ES, the throughput reaches 100 % of the available throughput (see figure 5).

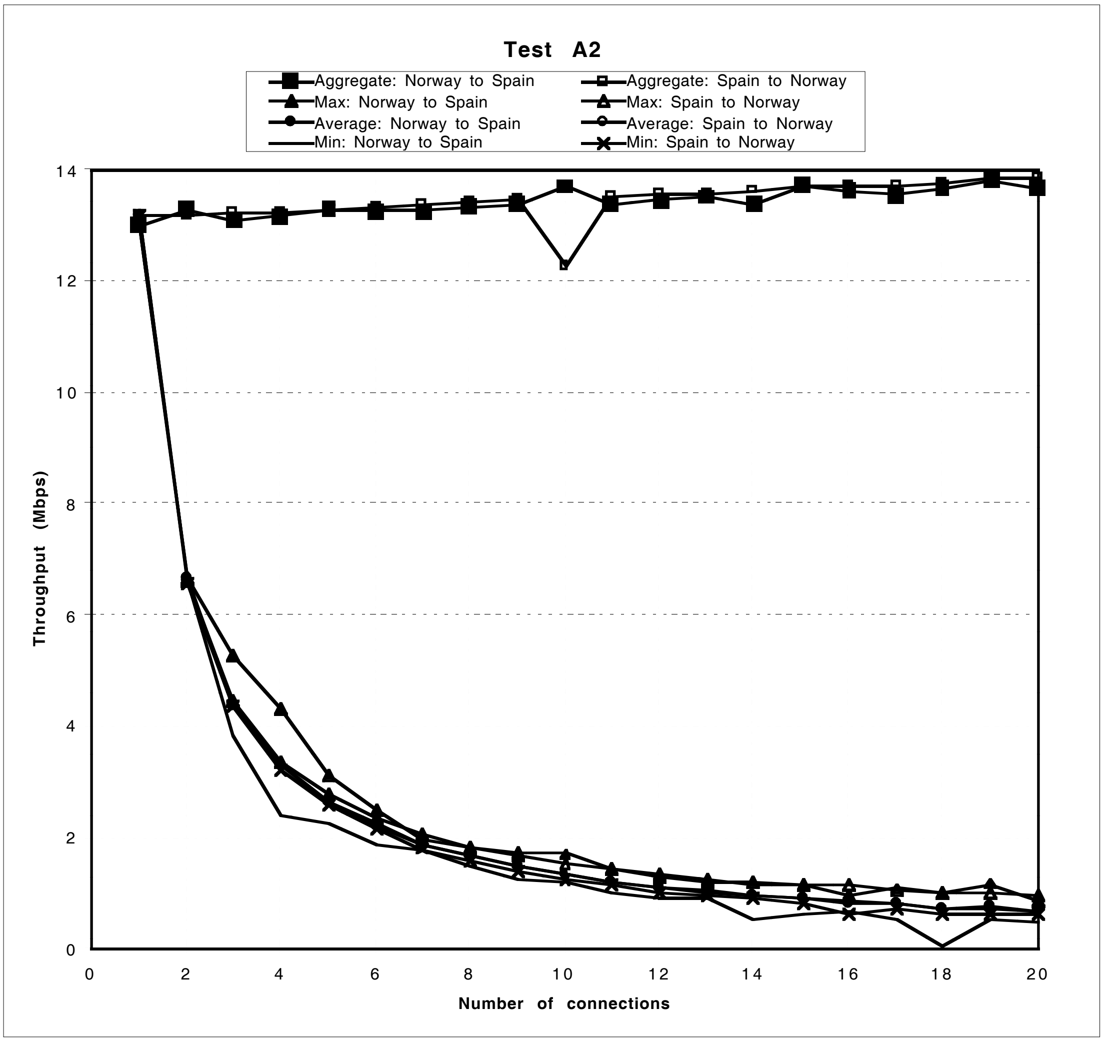

Figure 5: Test session Norway-Spain: Aggregated throughput for a bunch of one-way TCP connections between 1 pair of end-nodes.

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.

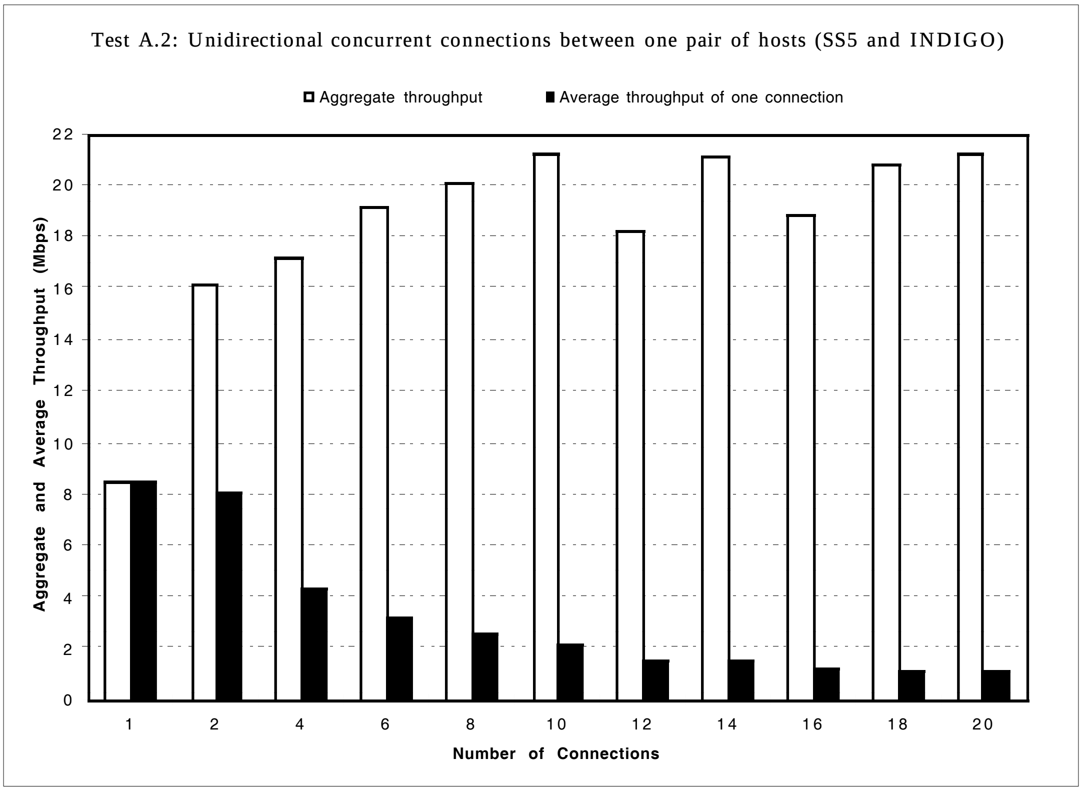

Figure 6: Test session Italy-Sweden: Aggregated throughput for a bunch of one-way TCP connections between 1 pair of end-nodes.

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.

2.2 Full duplex bandwidth level of occupancy

When concurrent connections are activated between two end-nodes 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-ES VP with about 13.8 Mbps, again, in both ways.

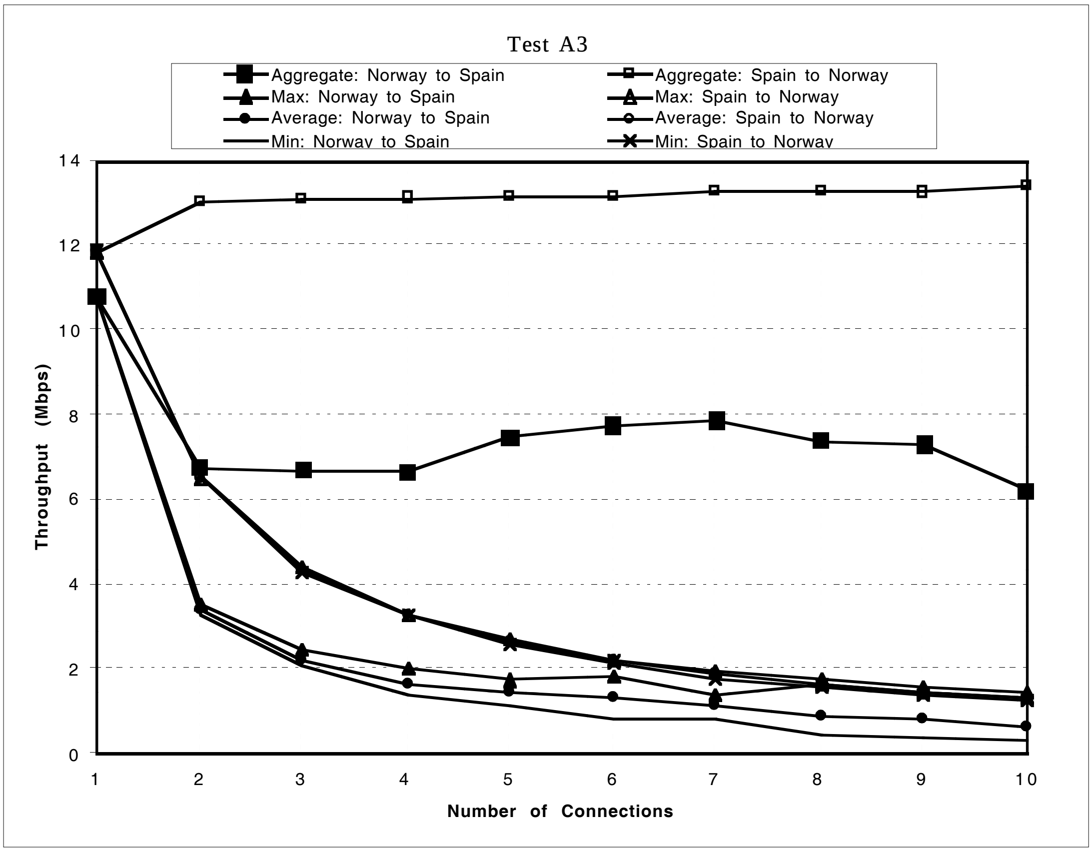

Figure 7: Test session Norway-Spain: Test of aggregate throughput for a bunch of two-way TCP connections between two end-nodes.

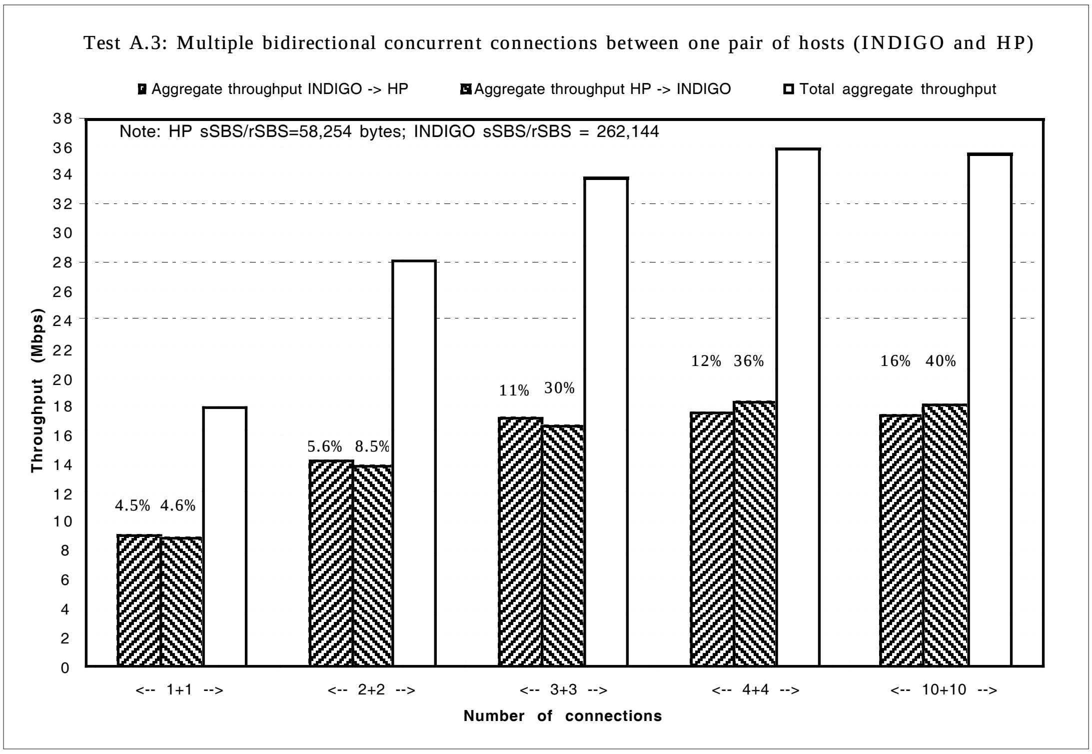

Figure 8: Test session Italy-Sweden: Test of aggregate throughput for a bunch of two-way TCP connections between two end-nodes.

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 is not clear and needs more study.

2.3 Peak cell rate configuration on VCC

When according to the configured traffic pattern more 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.

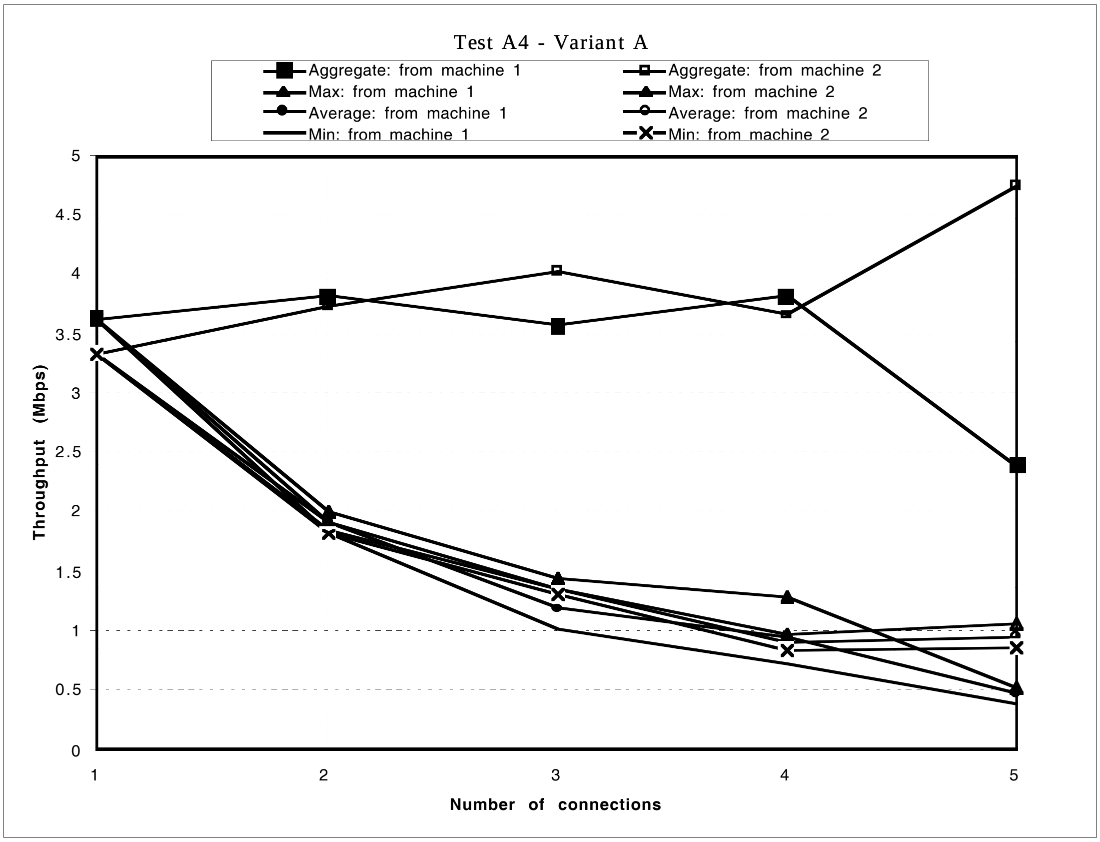

Figure 9: Test session Norway-Spain: aggregate throughput for 2 bunches of TCP connections from 2 senders to 1 receiver and with PCR =b on each VC.

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 pre-empted: 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 end-nodes, can't be overestimated:

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.

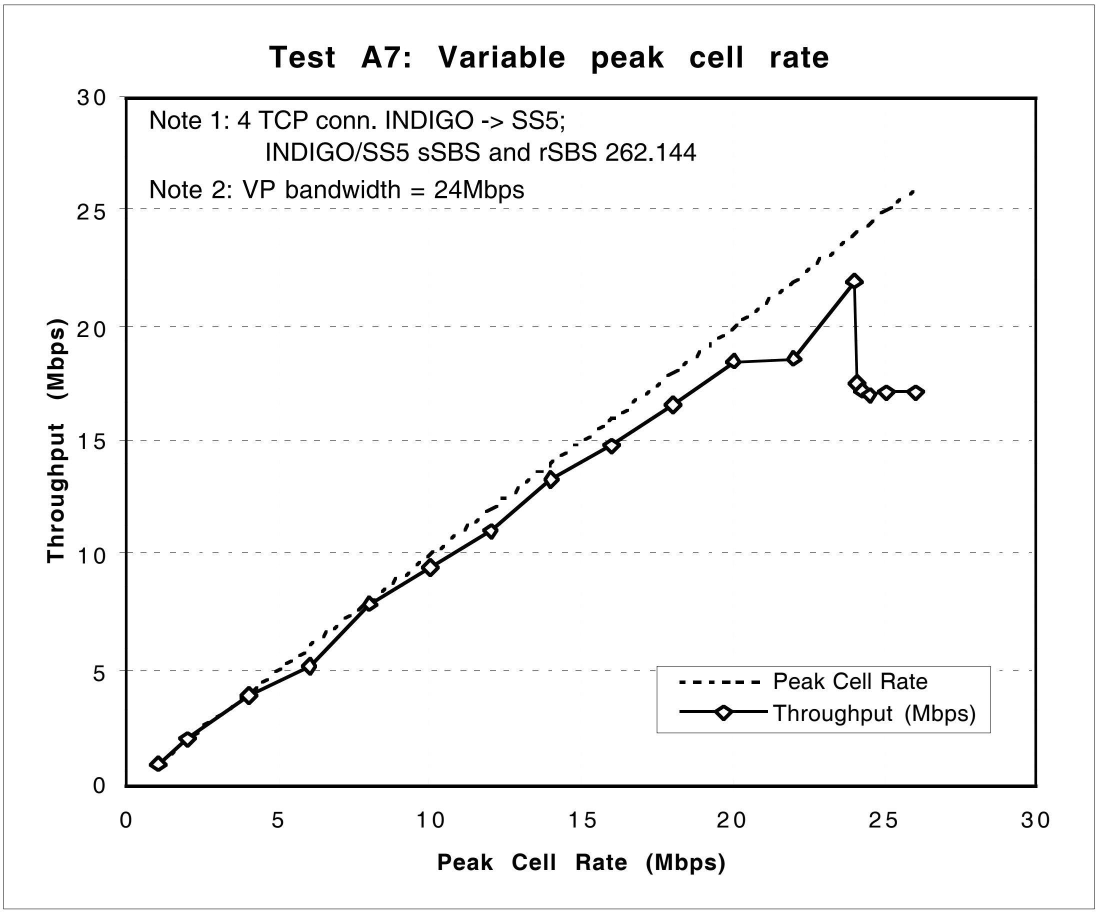

Figure 10: Relationship between throughput achieved by a one-way UDP connection for different datagram sizes and for different peak cell rate values assigned to the VC.

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 diagrams which are correct, i.e. can be sent down the network without being corrupted 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.

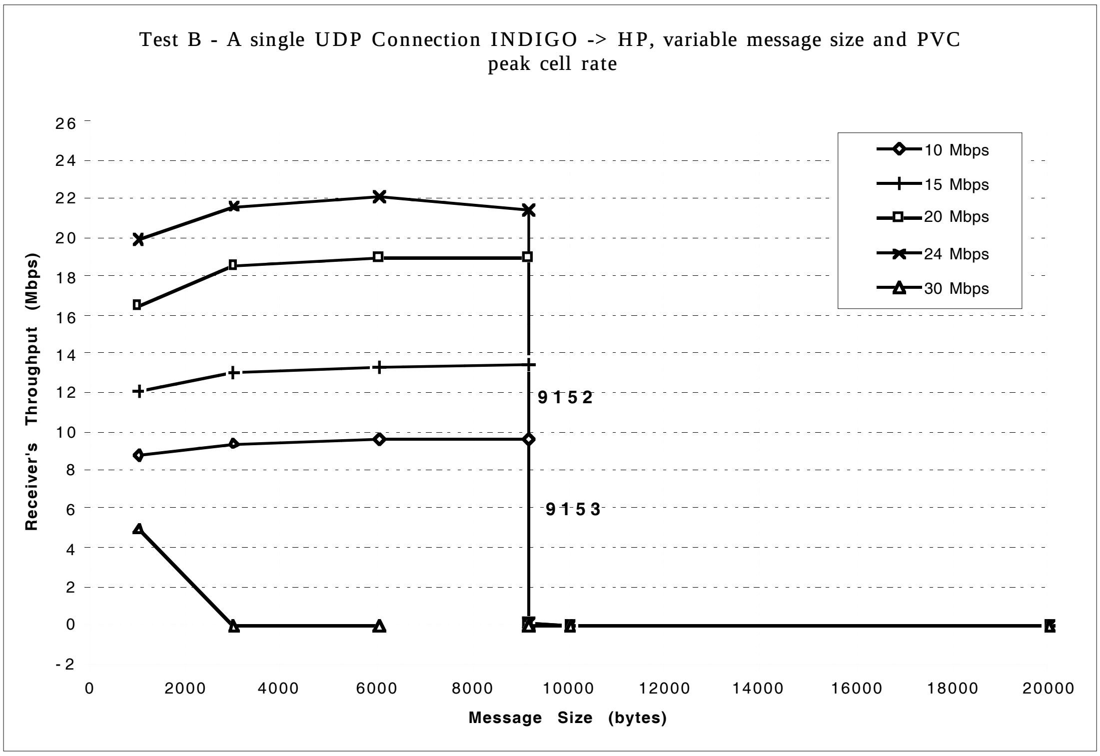

Figure 11: Relationship between variable peak cell rate (PCR) configured for the ATM VC and throughput achieved by a single one-way TCP connection on top of the VC itself.

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%.

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.

Test related problems and general comments

1. 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).

We know that ping command on some equipment only supports packets up to a given size. However this is not enough to explain the behaviour of end-systems in our experiment. In fact, in our case for the same sending platform the upper bound "B" of packet sizes depend on the current receiver "r" and this values B doesn't correspond to the maximum ping size of this receiver: with a different sender, receiver "r" can generate ping packets larger than B.

2. Cell drop with multiple one-way streams

As explained in the previous paragraph, if more connections are activated between two end-nodes, 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. The full understanding of this requires more investigation.

3. 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.

4. Inconsistency of politics 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.

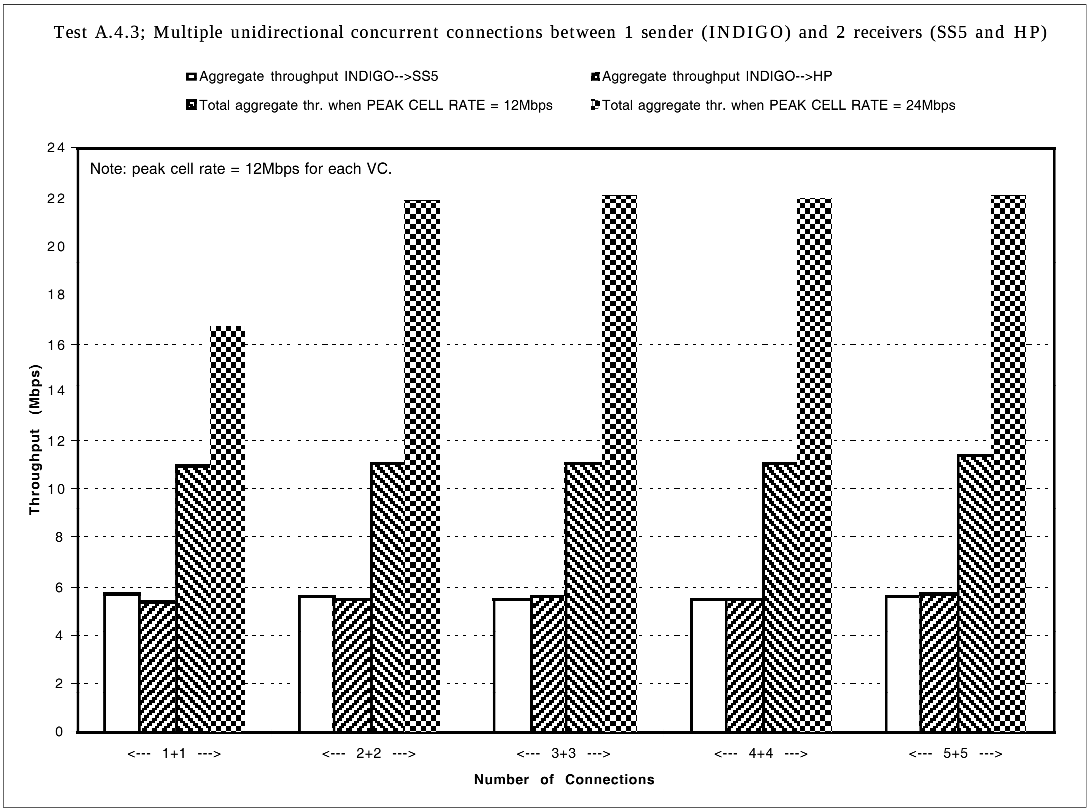

Figure 12: Comparison of aggregated throughputs achieved for 2 distinct bunches of one-way TCP connections from 1 sender to 2 receivers with different peak cell rate configurations.

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 are on the user equipment are required for a full understanding.

5. 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 resulting 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 end-nodes.

Further studies

The purpose of this test 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 show the relationship between degraded performance and VP length.

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 VPs 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.

5.2 SVC Tunnelling through PVPCs

Experiment leader

Christoph Graf, DANTE, Cambridge, UK

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 problematic. All equipment available to our tests opens best effort SVCs of traffic class UBR, without flow and congestion control. This works fine in uncongested LANs, but is completely unsuitable for operation across policed 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.

Interim solutions to this problem include the use of CBR instead of UBR for SVCs, or bandwidth limiting of (UBR) SVCs leaving a host. Using CBR instead of UBR produces serious problems with fair allocation between competing streams. The implementations we used support only UBR. Limiting the bandwidth of each SVC to a preconfigured value is available on some ATM adapters. But this creates another problem in return: you have to know the maximum number of SVCs beforehand and cannot allocate all bandwidth to a smaller number of SVCs. Configuring the maximum bandwidth of all SVCs leaving the host does not allow more than one host to use the same WAN link without risking to lose cells.

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 solve this problem.

Participants to the experiment

ACONET (AT):
- Gerald Hanusch, Universität Linz
- Guenther Schmittner, Universität Linz

INFN (IT):
- Mauro Campanella, INFN
- Diego Colombo, INFN
- Tiziana Ferrari, INFN/CNAF
- Simone Maggi, INFN

ULB (BE)
- Ramin Najmabadi, ULB

DFN (DE):
- Robert Stoy, RUS

SWITCH (CH):
- Simon Leinen, SWITCH

RESTENA (LU):
- Alain Frieden, RESTENA

Dates and phases

Phase one:	Set-up and test of local SVC infrastructure
Date:	August 96 - November 96 (individual to each participating site)
Duration:	approx. 3 weeks
Phase two:	Pairwise interconnection of participants over JAMES
Date:	August 96 - November 96 (individual to each pair of participating sites)
Duration:	approx. 2 weeks
Phase three:	Full interconnection of all participants over JAMES
Date:	started mid November 96
Duration:	3 weeks

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 in use:

VP	start date	end date	SVC up	JUD
AT-DE	19/08/96	09/09/96	29/08/96	tunnelling pre-tests
AT-DE	09/09/96	30/09/96	9/09/96	tunnelling phase 2
AT-DE	30/09/96	31/01/97	30/09/96	overlay network
AT-IT	20/09/96	30/09/96	-	tunnelling phase 2
AT-IT	30/09/96	31/01/97	03/10/96	overlay network
BE-DE	09/09/96	30/10/96	08/10/96	tunnelling phase 2
BE-DE	07/11/96	31/01/97	07/11/96	tunnelling phase 3
DE-LU	01/10/96	31/01/96	25/10/96	overlay network

Results and findings

Set-up of ATM equipment in all sites:

ACONET (AT):

set-up

```
PNO---GDC
    |
  LS1010
    |
  LS100--Linux
    |
  7010
```

ATM end systems

type and version: Cisco 7010
OS Version: 11.1(6)
ATM interface type and version: OC3 MM AIP
ATM driver version:

type and version: PC 486
OS Version: Linux
ATM interface type and version: OC3 MM by Efficient Networks
ATM driver version:

ATM switches

type and version: GDC APEX MAC XH
OS version: 4.1
ATM interface type and version: OC3 single mode/OC3 multi mode

type and version: Cisco LightStream 1010
OS version: 11.1(6)
ATM interface type and version: OC3 MM

type and version: Cisco LightStream 100
OS version: 3.1(2)
ATM interface type and version: TAXI MM

DFN/RUS (DE):

set-up

```
                +----(STM-1)--tencisco1 (193.246.0.55) (CISCO 7000)
                |
JAMES-(STM-1)-LS100---(STM-1)--tensun3   (193.246.0.52) (SUN SS 10)
                |
             (TAXI)
                |
             ASX200--(TAXI)---tensun4   (193.246.0.51) (SUN SS 10)
                |
             (TAXI)
                |
            LS100---(STM-1)--tensun2   (193.246.0.53) (SUN SS 20)
```

ATM end systems

router tencisco1 (193.246.0.55): CISCO 7010
SW-Version: IOS 11.1(5)

tensun2: 01
93.246.0.53
SUN SparcStation 20, Solaris 2.4
ATM-Adapter: HW Type: FORE sba-200e, media=oc3 multimodeVersion: 0.2.0
SW Version: Driver Version: A_ForeThought_4.0.2 (1.26)

tensun3: 193.246.0.52
SUN Sparcstation 10, Solaris 2.4
ATM-Adapter: HW Type: Efficient Network Inc. ENI-155s
SW Version: ATM Aruba ATM Software for Solaris 3.2.2

tensun4: 193.246.0.51
SUN Sparcstation 10, Solaris 2.5.1
ATM-Adapter: HW Type: Efficient Network Inc. ENI-100s
SW Version: ATM Aruba ATM Software for Solaris 3.2.2

ATM switches

Cisco Systems, Inc. LightStream 100
LS100 Software Version 3.1(2)
Boot ROM Version K02 22 Mar 95
ATMSIG Version 3.1(0.18) [960327:1035]
Switch ports:
OC-3c(SMF) connected to Cross Connect of Deutsche Telekom
OC-3c(MMF) connected to SUN Workstation and CISCO Router
100M-TAXI connected to FORE ASX 200

FORE ASX 200
Hardware version 1.0,
Software version ForeThought_3.4.3 (1.25)
Switch ports/network modules:
Output buffer size per port : 256 byte
HW-Type: NM-A-TAXI-100-4PT
HW-Version: A

INFN (IT):

set-up

```
JAMES--(E3)--ASX200--(STM-1)--Cisco7507
                  |   |
                  |   `---(STM-1)--Sgi
                  |
                  `------(STM-1)--Sun
```

ATM end systems

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(2),
RELEASE SOFTWARE (fc1)
AIP, hw 1.3, sw 20.06

SGI Indy, IRIX 5.3, GIA-200 adapter, 155 Mbit/s

ATM switches

FORE ASX-200 connected to JAMES at 34 Mbit/s (E3 port 1B1)
Hardware version 1.0, Software version ForeThought_3.4.0 (1.29)

RESTENA (LU):

set-up

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

ATM end systems

type and version: SS20
OS Version: Solaris 2.5
ATM interface type and version: Fore SBA-200 155Mbit/s
ATM driver version: ForeThought 4.0.0(1.30)

ATM switches

type and version: FORE ASX-200BX
OS version: ForeThought 4.0.0(1.30)
ATM interface type and version:

ULB/STC (BE):

set-up

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

ATM end systems

type and version: Sun SparcStation LX
OS Version: Solaris 2.5
ATM interface type and version: Fore SBA-200 155Mbit/s

ATM switches

type and version: Cisco LightStream 100
OS version: 3.1(2)
ATM interface type and version: STM-1

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 made 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 at this time, but it is expected to be caused by inefficient driver software. Furthermore, when establishing SVCs to tensun4 by use of 'ping', the first IP packet gets dropped.

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 under different load conditions

These tests are currently being carried out. Some preliminary tests were carried including both links IT-AT and AT-DE to set up SVCs between IT and
DE. Three intermediate switches were involved in setting up the SVCs:

Avg times (ms)	SUN-Cisco(DE)
SVC set-up	150.6
Next RTT Avg	56

The following results show set-up times between hosts in Italy and Brussels. They passed four intermediate switches involved in switching and three WAN links (IT-AT, AT-DE, DE-BE):

	SUN-CISCO(BE)	SGI-CISCO(BE)
SVC set-up	233.851	211.655
Next RTT Avg	86	86

The following major observations were made during the tests:

- Discrepancy between LAN and WAN: Even though the same equipment can be used to both interconnect local ATM hosts and WAN links, the configurations for each type of interconnect differ significantly. This is mainly due to the fact that WAN links require proper traffic shaping to cope with policed PVPs of much lower bandwidth than line speed in most cases. This is normally not required on a LAN due to the much higher available bandwidth. As a result, it may not be possible to use the same services over WAN links that can be used locally. The choice to use UBR for SVC makes it unusable over WAN links.

- End system limitations: The only application at hand to make use of our SVC infrastructure is the ATM/AAL5/IP stack. Unfortunately, it uses only UBR best effort SVCs and cannot be configured to use CBR/VBR SVCs.

- Limitations on switches: Even if end systems would be able to use other than UBR SVCs, i.e. CBR/VBR, not all switches currently support other SVCs than UBR. Another feature not generally available but required for reliable CBR SVC service is traffic reshaping and policing.

Implementation bugs detected so far:

- Cisco LS 100 switches need both ends of a VP hosting an SVC tunnel to use the same VPI number. The JAMES contacts were helpful to reassign VPI numbers to overcome this implementation limitation.

- Frequent crashes were reported when mainly SUN workstations with FORE cards should set-up SVC connections across Cisco LS100 switches. The exact set-ups and release versions of both hardware and software leading to this malfunction are currently under investigation.

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 connection between the same set of routers will be done by ABR SVCs resulting in a partial or full mesh. Depending of available services, either tunnelling or native SVC will be used. This set-up provides some major advantages in respect to the initial one:

- Switches introduce much 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 in the initial set-up.
- 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

Test-related problems and general comments

The JAMES procedures of setting up 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 the fact that ATM is known to be a complex technology, we almost always underestimated the effort required to setting up networks based on ATM.

Further studies

In the scope of this experiment:

- Conclude phase three testing, including testing of set-up time under increased load by parallel set-up requests.
- Trying to understand the high increase of set-up times over WAN links.
- The maximum packet size which can be transported across UBR SVCs including one or more WAN links vary significantly. This is currently not well understood but certainly linked to the CDVT/BT of JAMES VPs. Some collaboration between JAMES and the SVC testers is required to resolve this uncertainty.

Outside of the scope of this experiment:

- 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.
- Our tests are being carried out using signalling protocol UNI3.0 due to missing support of UNI3.1 on most of our equipment. Future test should rely on UNI3.1 instead.
- The only application tested so far was the ATM/AAL5/IP stack. Other applications should be considered as well.
- First implementations of the dynamic ATM routing protocol PNNI (phase 1) are available now and should be tested instead of the statically configured IISP used during our tests.

5.3 Classical IP and ARP over ATM

Experiment leader

Ramin Najmabadi Kia, ULB/STC

Summary of results

There was some initial testing of ARP in the context of the SVC tunnelling experiment (see 5.2). These initial tests revealed no major problems. The more generic test of ATMARP will be based on a SVC infrastructure which is the result of the SVC tunnelling through PVCs tests. This infrastructure will become available by the end of November.

Participants to the experiment

A subset of the SVC tunnelling test participants will work on the ATMARP tests. Details are to be confirmed.

Dates and phases

Phase one: Set-up and test of one Logical IP Subnetwork (LIS) with one ATMARP server located at the ULB/STC premises.
Start date: December 96
Duration: One week.

Second phase: Set-up and test of at least one other LIS (location to be determined).
Start date: Mid December 96
Duration: Two weeks

Network infrastructure

Mostly the same infrastructure used by the SVC tunnelling through PVCs experiment. ATMARP related software (server and client side) is required.

Results and findings

Some initial tests were made within the SVC experiment. Local ARP servers were set up within IT and AT, and accessed as in a normal LAN environment. This did not present any problems. Further tests about accessing ARP servers through the WAN, as well as the operations and interactions of several ARP servers still have to be carried out. These tests were delayed by the problems in the switching experiment.

Relevance for service and migration suggestions

For a switched ATM network an ARP service is essential to make it easily usable. It is too early to make statements about the usability of this technique in a production environment. The ARP is also only useful in conjunction with switching, so that the problems discovered there also hold back the deployment of the ARP service.

Test related problems and general comments

N/A.

Further studies

N/A.

5.4 IP routing over ATM with NHRP

Experiment leader
Olav Kvittem

Summary of results
The project is the investigation phase performing initial experiments. There are operative VPs to Stockholm and Madrid and SVC-tunnelling is being set up.

Dates and phases
The project is 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

Network infrastructure
Waiting for SVC-project to provide basics for a SVC-infrastructure over the overlay network.

Results and findings
An implementation is available on the Cisco routers
One contact with the IP on ATM responsible in JAMES awaiting feedback for collaboration.

Relevance for service and migration suggestions
Earliest use of the NHRP-service for application 97-02.

Test related problems and general comments
There are some initial problems getting ATM SVCs to work.

Further studies
N/A

5.5 European ATM Addressing

Project Manager

Kevin Meynell, UKERNA

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.

Participants

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

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.

Further studies

It is necessary to continue to monitor progress on the standards relating to addressing at the ATM Forum and the ITU.

5.6 ATM Network Management

Experiment leader:

Zlatica Cekro, University of Brussels, ULB/STC

Summary of results

Experience is gained on ATM user interface and some elements of end-to-end management. A private management platform based on SNMPv1 and SNMPv2 standards has been applied on a subset of NRN ATM segments for “monitoring information” i.e. read only class of service. A transport link between the management platform and NRN ATM devices was realised through the operational Internet service with access to an ATM switch or a router through an Internet address. Test results concern the following:

- SNMP is widely in use in the edge devices: routers with ATM interfaces and ATM switches;
- Successfully tested were SNMPv1 protocol based: IETF Interface MIB in MIB-II, Cisco LS100 proprietary ATM MIB, Cisco router proprietary MIBs and ATM Forum UNI MIB.

NRNs ATM equipment participated in the tests include:

- NORDUnet: Norway, Oslo, Cisco ATM switch LS100
- GARR: Italy, Milan, FORE ATM switch ASX200
- ULB/STC: Belgium, Brussels, Cisco ATM switch LS100 and Cisco router 7010.

Management platform was SunNet Manager-SunNet Domain Manager version 2.3 on Solaris 2.4.

Participants

For Phase 1 and 6 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 other phases (for the moment): NORDUnet (SE and NO), GARR (IT), ULB/STC (BE).

Dates et phases

Phases are the same as in D 11 except that dates are moved in + for about one month.

Network infrastructure

No special infrastructure was requested except the support of MIB II, ILMI (Integrated Layer Management Interface) UNI MIB for the UNI at the user portion of the Public ATM network.

Local infrastructure

No special infrastructure was required except a Management system station with SNMPv1 and SNMPv2 for monitoring the specific view and data collecting.

Hardware/software

The releases of software that support the latest standards are required like MIB II and SNMPv2 and ATM Forum MIBs.

Results and findings

Results of phase 1:

The work has been realised in July and August 1996. Investigation of management possibilities on the NRN and PNOs side resulted in modification of our test scenario, Version 2 from July 1996. The new scenario, Version 3 from September 1996, includes the tests of the same management possibilities as in the original version but applied only at the ATM NRN edge devices which participate in the TEN-34 experiments. Instead to base the tests on M3 interface - Customer Network Management for ATM Public Network Service (which includes an active role 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.

The new scenario includes two items:

1 Usage of a Management platform based on the SNMP with a private view of the NRN ATM infrastructure. Realisation of the tests with the M2/M3 interface should be according the test scenario Version 3, phases 1-6. The Management platform should include the ATM switches from different vendors and should be based on SNMPv1 and SNMPv2 protocols. The platform is planed to be shared by the participants in the tests by means of remote access.

2 Creation of a WWW based TEN-34 management page for public access with a subset of relevant views and statistics of interests for the TEN-34 experiments.

Results of Phase 2, test scenario Version 3:
The work has been done during September 1996. Access to the edge ATM devices was realised 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. If the firewall is not applied it is possible to have available different SNMP management functions depending on the user privileges (public or private privileges). The SunNet Management platform from ULB/STC premises (SunNet Domain Manager version 2.3 on Solaris 2.4) is used for the access to the MIBs in three different locations:
- NORDUnet: Norway, Oslo, Cisco ATM switch LS100, Software version 3.1(2)
- GARR: Italy, Milan, FORE ATM switch ASX200, Software version ForeThought 3.4.0(1.29)
- ULB/STC: Belgium, Brussels, Cisco ATM switch LS100 and Cisco router 7010, Software version IOS 11.1(4).

Results of Phase 3, test scenario Version 3:
The work has been done during the October and the first half of November 1996.

The tests included public "read only" management functions:
- SNMPv1 IETF MIB-II read access and statistics collecting
- SNMPv1 read access to LS100 proprietary ATM MIB
- SNMPv1 Cisco router proprietary MIBs
- SNMPv1 ATM Forum UNI MIB.

Through the remote access (X window terminal) all participants in the tests could benefit of usage of the SunNet Manager console as it allows multiple users to work simultaneously.

The ATM management functions were tested locally by ULB/STC and remotely by SURFnet.

Phases 4-6 adapted to the test scenario Version 3 will be continued as planned.

Test related problems and general suggestions
The test scenario has been changed from the tests on public to the private ATM segments due to the lack of the general management services offered by PNOs (JAMES). The modifications of the test scenario assume existence of similar functionality on the edge NRN devices and on the public segments which correspond to the specific user. Other specific functions like "OAM end-to-end" has to be asked explicitly from the PNOs through a JUD (JAMES User Document).

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). CMIP based management platforms with X.user interface could offer more complex functionality but the process is still in the standardisation phase.

Test-related problems and general comments

More ATM switches from different vendor should be included (like GDC, UB). The lack of standardised tools on the WWW based management provoke the delay in the realisation of TEN-34 Network Management Web page.

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.

5.7 CDV over concatenated ATM networks

Experiment leader

Victor Reijs

Summary of results

Initial tests have been started locally and procedures are being tuned and bugs in the procedures ironed out. We have started preparations for the international tests.

Participants to the experiment

SURFnet b.v. (The Netherlands), University of Twente (The Netherlands), University of Stuttgart (Germany), UNINETT (Norway), JAMES

Dates and phases

We plan to start the international measurements in the first two weeks of December, 1996. Starting in November 1996, we began the local phase of the experiments and that continues through December. There will be an analysis phase in January 1997.

Goals of the project

The goal of this project is to find out how various cell streams change as they pass through an ATM network. This information is useful (essential) to know when traffic characteristics have to be negotiated for individual connections between administrations or service providers.

The plan is to capture the cell stream characteristics both before and after the stream traverses the JAMES network. For this, we need measurement equipment at each access point to the JAMES network. We will look at both CBR and VBR cell streams; and we will generate these cell streams both in isolation and also multiplex them (at the first switch after the source) with other cell streams in order to try to understand how other traffic, as well as conditions in the networks, might affect the cell stream characteristics. The measurements that we want to take consist of traces of the cell stream(s) of interest. These cell stream traces will then be analysed in order to 'fit' a set of traffic characteristics to them.

Network infrastructure

Contact has been made with people from JAMES to arrange for the facilities necessary to execute the international portion of the project. Dirk Hetzer of DT Berkom has started this process and will contact the PTTs of the Netherlands and Norway, as well as people in the German network, to try to arrange for the permanent virtual circuits and the measurement equipment necessary to carry out the experiments. In addition, the participants in the individual countries must set up PVCs also from their individual sites to the national host in each country and thence to JAMES. The proposed participants are the University of Twente (the Netherlands), the University of Stuttgart (Germany) and UNINETT (Norway).We will need 2 PVCs per destination for the test; that is, 2 for Norway and 2 for Germany. For the isolated cell stream tests, only 1 PVC will be used and for the multiplexed cell stream tests, both PVCs will be used. DT Berkom has an HP 75000 analyser that they can use to measure the cell stream at the exit of the JAMES network in Germany.

Results and findings

Local testing has started at the University of Twente to test the traffic generation program and the measurement equipment. We are using a SUN Ultra 1 with a SUN ATM 155 card as the main traffic source. For measurement equipment we are using a RADCOM RC-200, which is soon to be replaced with a RADCOM PRISM 200 (formerly the ATMmax). So far we have tested CBR cell streams at 64 kbps and 128 kbps (due mostly to limitations in the RADCOM's capture speed and time-stamping resolution; we expect these problems to be relieved somewhat with the new RADCOM). The current RADCOM has a time-stamp resolution of 1 microsecond and a capture memory of about 22000 cells if you capture only the cell header (which is all that is really necessary in this experiment).

Initial results from these local tests have shown that some program characteristics and some test environment characteristics can introduce a wide variation in the inter-cell times of even a 'CBR' cell stream. We are currently working to eliminate these variations from the traffic generation process and produce a reasonably smooth cell stream for CBR. Our experience so far leads us to believe that we can eliminate most of this undesirable behaviour.

The international phase of the project will be carried out in December 96, with cell streams being sent to the other participants and measurements being collected at as many points as possible and processed by the University of Twente. The raw data (the traces) and the results will, of course, be shared with all participants.

Relevance for service and migration suggestions

As long as there is no cell shaping at every boundary system, alignment of CDVT/BT should be done within every domain.

Test related problems and general comments

Getting the right test equipment with enough accuracy for determining cell delay variations.

Further studies

None so far

References

(http://wwwtios.cs.utwente.nl/~chimento/tften/tftenpage-1.html) user id (tften) and password (tftenifsp)

5.8 Performance of the Native ATM Protocol

Experiment leaders

Mauro Campanella, INFN (Milan) and Tiziana Ferrari, INFN/CNAF (Bologna)

Summary of results

Goals of the test are the monitoring of the native ATM protocol performance in the infrastructure of JAMES through the measurement of some network parameters and the analysis of the network behaviour (fairness of bandwidth distribution, maximum achievable throughput) when the infrastructure is stressed by different patterns of traffic.

The initial phase of this test will start when results of tests of performance of TCP/IP over ATM will be completely collected and results fully analysed. In fact, one of the targets of this experiment is to compare network performances with and without the TCP/IP layer implemented on top ATM and for this reason, test of TCP-UDP/IP performance over ATM and this experiment are strictly connected.

More work is planned for this test to install and analyse "TCP ONIP" ("TCP over Non-Existent IP"), in conjunction with the research done in this field at CNET (France) (see [1]).

Network infrastructure

There are no changes from the requirements laid out in the specification of the experiment in deliverable D11.1.

Results and findings

N/A

Relevance for service and migration suggestions

Through a comparison between the TCP/IP/ATM protocol stack and native ATM performance we can measure the impact of the TCP flow control mechanism over the maximum throughput achievable by an application generating a memory-to-memory data transfer and the impact of cell drop in both cases. In fact, when TCP/IP is implemented on top of ATM, in case of cell loss, all the corrupted packets are re-transmitted as an effect of the TCP protocol reliability, while, with native ATM, cell drop is not recovered and no re-transmissions occur.

Finally, through direct comparison of the numbers of sent and received ATM cells it is possible to measure the precise number of cells lost in the global JAMES infrastructure.

Test related problems and general comments

N/A

Further studies

N/A

5.9 Assessment of ATM/VBR class of service

Experiment leader

Olivier Martin

Summary of results

Some national testing was carried out in the Netherlands. These tests have revealed that VBR can be used to transmit general IP traffic, but extensive fine-tuning is required to make optimal use of the facilities. These tests have to be verified in the wide area network, where nothing has been done yet, as there is no VBR service available over JAMES yet.

Network infrastructure

There are no changes from the requirements laid out in the specification of the experiment in deliverable D11.1.

Results and findings

There are national measurements available (see http://www.nic.surfnet.nl/surfnet/persons/reijs/sn4/pcr.htm), which need to be verified in an international environment. All goals from the specification of the experiment are outstanding.

Relevance for service and migration suggestions

N/A

Test related problems and general comments

N/A

Further studies

N/A

5.10 IP resource reservation over ATM

Experiment Leader

Olav Kvittem, UNINETT, Norway

Summary of results

The project is in the investigation phase performing initial experiments. University of Dresden has developed RSVP over IP over ATM for DEC Workstations so that this functionality could be tested in a local ATM environment using an own performance tool with an integrated graphical RSVP user interface. Moreover, the U of Dresden is developing a video conferencing system testing RSVP over ATM by a more practical-relevant example.

Further participants

Frank Breiter, Sabine Kühn, Technical University of Dresden, Germany

Dates and phases

The project is 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 a SVC overlay infrastructure.

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

Network infrastructure

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

Our experimental environment consists of several multimedia workstations of type DECstation 3000 AXP 700 and 300 which are connected with a DEC Gigaswitch/ATM via multimode fibre. The cell transmission is performed using SONET/SDH frames with standard 155 Mbps per channel. Only AAL5 is currently implemented. The switch and ATM adapter cards support UNI 3.0 signalling as well as UNI 3.1 and offer NSAP/E.164 addressing. Moreover, CBR (constant bit rate), and ABR (available bit rate), both with point-to-point and point-to-multipoint VCs, are possible. Our local environment is currently 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-infrastructure over the overlay network for performing the pilot experiment.

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 new VC?).

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 ATM shortcuts 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.

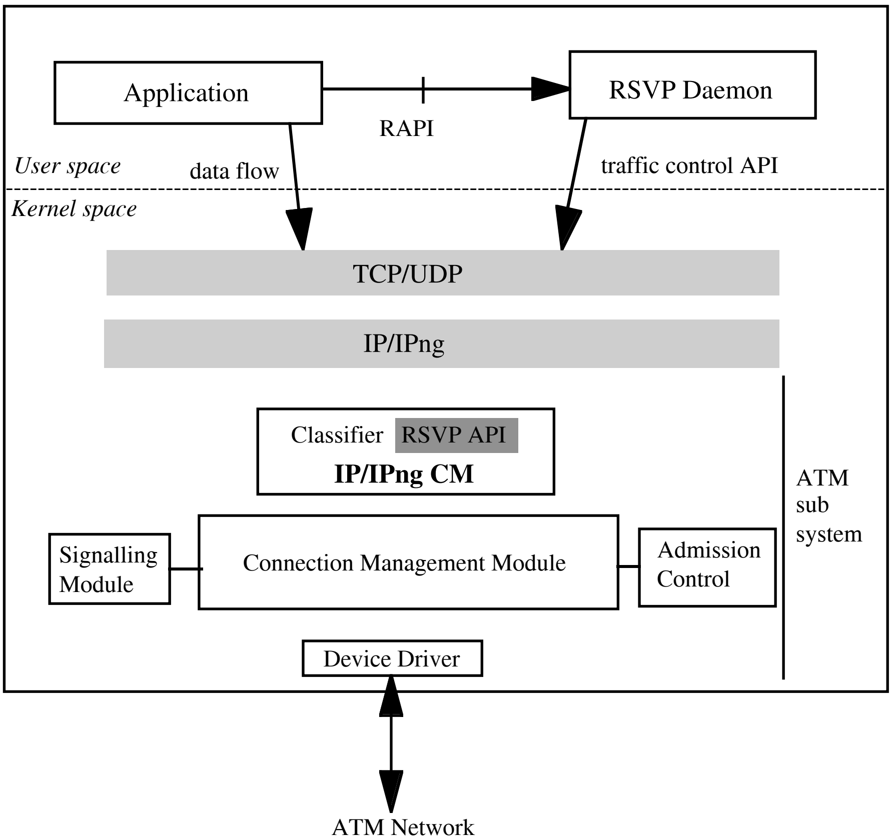

Figure 1: Interaction between ATM and RSVP (DU)

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 real time handling and transport of IP flows (figure 1). The API is also able to receive reservation information from a local daemon as a part of 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. A classification 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

Currently only unicast test 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. 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 experiments could be performed.

In preparation to 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.

Currently the tests ate planned between Norway and Germany. DEC equipment are available in Norway as we are not know about other existent RSVP over ATM implementations.

Test related problems and general comments

There are some initial problems getting ATM SVCs to work. Moreover, currently only the University of Dresden has a public ATM connection (DFN: B-WiN) not the Faculty of Computer Science itself. How could we get a JAMES admission with our DEC equipment?

Further studies

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

Experiment leaders
Paulo Neves and Roberto Canada

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

Participants to the experiment
Not yet defined.

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

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.

Results and findings
The initial phase of the work package 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 re-transmission 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 VIP/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 authentication integrity relay detection padding (against traffic flow confidentiality attacks)	authentication integrity replay detection	confidentiality integrity replay detection

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.

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.

Further studies

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

a) Masquerade;
b) Protocol spoofing;
c) Denial of Service;
d) 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.

Bibliography and References

Experiment 1

[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).

Experiment 2

[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

Experiment 3

Basic Documents:
[1] Classical IP and ARP over ATM (RFC 1577)
[2] Multiprotocol Encapsulation over ATM Adaptation Layer 5 (RFC 1483)
[3] Default IP MTU for use over ATM AAL5 (RFC 1626)

Additional Information:
[4] Integration of Real-time Services in an IP-ATM Network Architecture (RFC 1821)
[5] ATM Signaling Support for IP over ATM (RFC 1755)
[6] IPng Support for ATM Services (RFC 1680)
[7] NBMA Address Resolution Protocol (NARP) (RFC 1735)
[8] ASSIGNED NUMBERS (RFC 1700)
[9] INTERNET NUMBERS (RFC 1166)

Experiment 4

[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

Experiment 5

[1] SVC Tests
[2] ARP Tests
[3] Howat, G; JANET ATM Addressing Scheme; UKERNA; 1996
[4] Howat, G; ATM Addressing Discussion Paper; University of Edinburgh; 1996
[5] ATM Forum Technical Committee; ATM Forum UNI 3.1 Specification; ATM Forum; 1994
[6] ATM Forum Technical Committee; ATM Forum UNI 4.0 Specification; ATM Forum; 1996
[7] ATM Forum Technical Committee; ATM Forum PNNI 1.0 Specification; ATM Forum; 1996

Experiment 6

[1] IETF RFC 1695, M. Ahmed and K. Tesink, "Definitions of Managed Objects for ATM Management", Version 8.0, August 1994
[2] IETF RFC 1442, J. Case, K. McCloghrie, M. Rose, S. Waldbusser, "Structure of Management Information for version 2 of the Simple Management Protocol (SNMPv2", May 1993
[3] IETF RFC 1336, J. Galvin, K. McCloghrie, "Security Protocols for version 2 of the Simple Network Management Protocol (SNMPv2), May 1993
[4] ITU-T, M.3020, Maintenance, Telecommunications Management Network, TMN Interface Specification Methodology, July 1995
[5] ITU-T, M.3100, Maintenance, Telecommunications Management Network, Generic Network Information Model, July 1995.

Experiment 7

Experiment 8

[1] A Platform for TCP over Native ATM; Afifi Hossam, Bonjour Dominique, Elloumi Omar, Proceedings JENC7, paper 153; 1996.
[2] Permanent virtual circuits configuration and TCP/UDP/IP performance in a local ATM network; C.Battista, M.Campanella, T.Ferrari, A.Ghiselli, C.Vistoli, INFN Internal Note n.1069, July 1995..
[3] Performance evaluation of TCP(UDP)/IP over ATM networks; S.Dharanikota, K.maly, C.M.Overstreet, Computer Science Dep., Old Dominion University, Norfolk VA.
[4] 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
[5] High Performance TCP in ANSNET; C.Villamizar, C.Song.
[6] High-performance TCP/IP and UDP/IP Networking in DEC OSF/1 for Alpha AXP Digital Technical Journal, vol. 5, n. 1, win 1993.
[7] How a large ATM MTU causes deadlocks in TCP data transfers; K.Moldeklev, P.Gunninberg (Norwegian Telecom Research and Swedish Institute of Computer Science).

Experiment 10

Project plan : via http:/www.dante.net/ten-34/tf-ten

Experiment 11

[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 securité 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.

Glossary

ARP	Address Resolution Protocol
ATM	Asynchronous Transfer Mode
CBR	Continuous BitRate (ATM Forum: traffic class)
DBR	Deterministic BitRate (ITU-T: traffic class, eq CBR)
E.164	(ITU-T addressing standard)
ILMI	Interim Link Management Interface
IP	Internet Protocol
LIS	Logical IP Subnetwork
MBS	Maximum Burst Size (ATM Forum: traffic parameter)
NHRP	Next Hop Resolution Protocol
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
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