



**Project Number: RE 1009**

**Project Title: TEN-34**

## **Deliverable D14.2 Results of Phase 2 Test Programme**

Deliverable Type: PU - Public  
Contractual Date: 30 April 1998  
Actual Date: -  
Work Package: 14 - Phase 2 Test Programme  
Nature of Deliverable: RE - Report

### **Authors:**

Stefania Alborghetti, INFN, IT	Simon Leinen, SWITCH, CH
Michael Behringer, DANTE, UK (editor)	Kevin Meynell, TERENA, NL
Zlatica Cekro, VUB/ULB, BE	Victor Reijs, SURFnet, NL
Vegard Engen, UNINETT (BDC), NO	Günther Schmittner, Johannes Kepler University, AT
Tiziana Ferrari, INFN/CNAF, IT	Robert Stoy, RUS, DE
Christoph Graf, DANTE, UK	Jean-Marc Uzé, RENATER, FR
Olav Kvittem, UNINETT, NO	José Vilela, FCCN, PT

### **Abstract:**

*Workpackage 11 of the TEN-34 carried out the first phase of the TEN-34 experimentation programme. Workpackage 14, the second phase of the test programme, deals with more complex experiments such as ATM routing, and advanced some of the experiments from phase 1 further. In deliverable D14.1 the experiments of this phase were defined and this deliverable, D14.2 describes the results. D14.3 will provide a summary of the complete TEN-34 testing programme.*

*The experiments of this phase cover a wide range of wide area networking technologies such as ATM routing and signalling, and related issues like call admission control. Technologies to map IP services onto ATM were also discussed, such as label based switching and NHRP.*

*As in phase 1 the experiments were carried out over the JAMES network, which will be described in*

*more detail in this deliverable. The JAMES network provided a European ATM network facility, and the TF-TEN used an overlay network consisting of VPs on this network for most of their experiments.*

*The results of this workpackage show that the basic technologies to use ATM services and high-level IP services are working in a test environment. However, the mechanisms to control and manage these advanced technologies are not fully developed yet, which makes the technologies unusable in a production network. New ATM features such as point-to-multipoint SVCs are becoming available, but need more investigation. The integration of IP and ATM has made significant progress over the last year with the development of label based switching technology.*

**Keywords:**

*ATM, TEN-34, TF-TEN, JAMES, VP tunneling, ATM routing, PNNI, label based switching, tag switching, ATM resource reservation, IP resource reservation, RSVP, ATM point-to-multipoint, ATM signalling, SVC, UNI, ATM policy control, CAC, ATM accounting, ABR, ABR RR, ABR EFCI, ABR ER, NHRP, ATMARP, ATM addressing, NSAP, E.164, ATM address translation, native ATM performance, TCP ONIP, Arequipa, ATM API, ATM network management, OAM, SNMP, ATM security*

## **[Table of Contents](#)**

---

### **Other formats**

[D14.2 version 980611v1: [Word 8](#) (2.2 Mbyte), [Word 8, gzip](#) (0.8 Mbyte) ]

# D14.2 - Table of Contents

---

## [Executive Summary](#)

1. [Summary of Results per Experiment](#)
2. [Usage of the JAMES Network](#)
3. [Conclusions](#)
4. [Detailed Test Descriptions](#)
  - 4.1 [ATM Routing](#)
  - 4.2 [Label based switching](#)
  - 4.3 [ATM resource reservation](#)
  - 4.4 [IP resource reservation](#)
  - 4.5 [ATM point-to-multipoint](#)
  - 4.6 [ATM signalling](#)
  - 4.7 [ATM policy control and accounting](#)
  - 4.8 [ATM traffic management](#)
  - 4.9 [ATM address resolution](#)
  - 4.10 [ATM addressing](#)
  - 4.11 [Native ATM performance](#)
  - 4.12 [ATM network management](#)
  - 4.13 [Security in ATM Networks](#)

## [Glossary](#)

---

[Back](#) to cover page

## 4.13. Security in ATM Networks

### Experiment leader

José Vieira Vilela, FCCN, Lisbon, Portugal

### Summary of results

True WAN ATM networks can either be built with a subset of the ATM functionality - CBR PVCs, no signalling support, no QoS - or with the full capabilities of this technology. In the first case, the security issues involved are more or less the same as when using leased-lines. However, when you plan to deploy fully functional Wide Area ATM networks, with support for truly advanced services through negotiated call establishment, security concerns become much more important. Intermediate switches, public and private, must support signalling, exchange information, and act accordingly. It's at this point that the switches, and the network itself, become vulnerable to several threats. And the implementation of specially designed security mechanisms in ATM networks becomes an unavoidable necessity.

In this work, the state of the ongoing specification effort on ATM security issues, and the implementation of these specifications and overall security features on two of the most popular ATM switches available are analysed and described.

### Participants to the experiment

Rui Bettencourt, FCCN, Portugal.

### Dates and Phases

**Phase 1:** Review of the security-related specifications of the main International Standardization bodies

Date: July 97 - March 98

**Phase 2:** Survey on the implementation status of security-related standards on current ATM equipment

Date: September 98 - March 98

---

## Results and findings

### PHASE I

The standardization bodies involved in ATM specification started working on security-related issues specific to ATM networks very recently.

On its X Series, the ITU-T produced general security recommendations for data networks and open systems communication (see X.800 - X.835) and some ATM specific security issues are addressed by isolated documents of the I Series (I.251, I.255, I.259) and the M Series but there is, to our knowledge, no effort going on to produce a systematic set of standards for ATM security.

In October 1995, the Technical Committee of the ATM Forum created a specific Working Group to "identify security problems resulting from the technology and investigate solutions to those problems". The works of this group suffered, however, many delays, to the point that at the time of this writing, there are still no ATM Forum approved specifications on security. The "ATM Security Specification Version 1" is now in Straw Ballot (which will finish in April) and it has expected final approval in July 1998. The ATM Security Framework 1.0 was presented for Final Ballot last January and may still be approved this year, and a "Security Addendum" to the UNI Signalling 4.0 is also due for approval in December 1998. A review on the main points addressed by these documents being discussed by The ATM Forum seems still useful, since their structure and objectives is not likely to change until final approval.

The implications of ATM security issues in protocols operating over it (e.g. IP) is of special importance for NRNs. When IP is carried over PVCs the difference from the common leased-lines transmission is negligible, but on what Advanced Services are concerned, we are implicitly counting on ATM signalling. Several IETF standards have already been released that take advantage of signalling to provide truly advanced services in IP networks [11, 12] and their security requirements and problems are also being analysed by the IETF [13, 14].

## IETF Work

Basically, three upper layer protocols are examined by the Internetworking over NBMA (ION) working group: **CLIP** [8], **MARS** [11], and **NHRP** [12]. These protocols rely on UNI 3.0/3.1 signalling to provide basic IP connectivity between host on the same Logical IP Sub-network (CLIP), intra-subnet multicast connectivity (MARS), and direct (shortcut) connectivity between hosts and routers in different subnets (NHRP).

All the security problems detected in these technologies derive from two basic ATM security weaknesses: the lack of end-node authentication mechanisms, and the lack of access control. In all cases, the possibility for a host to present a bogus address in the Calling Address Information Element of the SETUP message is the hardest problem to solve, since it implies the possibility of IP address spoofing and simplest to implement upper layer access control schemes are based on this address.

ATMARP servers are completely vulnerable to database poisoning by an attacker and allow the scanning of an entire LIS; Impersonation of the ATMARP server is also possible in some cases.

In MARS (maybe the most vulnerable of all three), it is possible to join a cluster and receive all traffic on any active multicast group simply by knowing the NSAP address of the Multicast Address Resolution

Servers (MARS). Several kinds of DoS attacks - to the address server, to the Multicast Server (MCS), and to clients - are possible and require little *a priori* knowledge of the cloud.

NHRP has some provision for upper-level authentication that may preclude the same kind of threats that exist to the above protocols, but if this option is not implemented it may become the bigger source of problems, since a Next Hop Server (NHS) essentially accepts calls from any address (in the same or in other LIS), and is visible from the entire Internet (with traceroute, for instance). Together with MARServers, NHS are also the easiest entities to overload by repeated requests.

ATM Forum standards MPOA and LANE suffer from essentially the same weaknesses as the above mentioned protocols. In both cases there are servers that accept calls from clients, register the information they have to offer, and provide them with the data they request (namely, addresses of other clients). In LANEv2, an elementary access control mechanism (based on the callers NSAP address) may optionally be implemented at the LECS, in MPOA there isn't any provision for access control whatsoever.

## ATM Forum Work ATM Security Framework 1.0

The aim of this document is to identify generic security requirements for security services provided within ATM networks and to derive their basic supporting services. It provides a theoretical analysis on the security objectives of the interested parties (customers, operators, and public authorities), the threats to these objectives, the resulting security requirements to counteract these threats, and an ATM terminology formulation for these requirements. The security services that shall fulfill the functional requirements are listed and explained. The plane specific interpretation of the functional security requirements is partly described (User Plane only).

- **Generic security objectives**
  - Confidentiality
    - of stored and transferred information
  - Data Integrity
    - of stored and transferred information
  - Accountability
    - for all ATM network service invocations and management activities
  - Availability

of the ATM facilities to the legitimate entities

- Generic **threats** to these objectives

- Masquerade, Impersonation, or Spoofing

An entity pretends to be another

- Eavesdropping

Communication interception or monitoring

- Unauthorized access

An entity attempts to access data in violation of the security policy in force

- Loss or Corruption of Information

Data in transit is modified, by insertion, deletion, replay, delay or replacement of segments or complete messages.

Stored data (for example, billing data) is modified or deleted

- Repudiation

An entity involved in a communication exchange subsequently denies the fact.

- Forgery

An entity fabricates information and claims that this information was received or sent from/to another entity

- Denial of Service

Whenever a entity prevents other entities from performing their functions

The following table summarizes the mapping between the security objectives and the threats.

Objectives	Generic Threats						
	Masquerade	Eavesdropping	Unauthorized Access	Corruption or Loss of Data	Repudiation	Forgery	Denial of Service
Confidentiality	yes	yes	yes	--	--	--	--

Data Integrity	yes	--	yes	yes	--	yes	--
Accountability	yes	--	yes	--	yes	yes	--
Availability	yes	--	yes	yes	--	--	yes

- Functional security **requirements** to counteract the security threats

- Verification of identities

- Controlled Access and Authorization

Implies verification of identity and applies to objects such as the physical systems, their software, and data

- Protection of Confidentiality

of the user related ATM network information (e.g.. billing information and data exchanged in private sessions), and information used to support security services (like cryptographic keys)

- Protection of Data Integrity

of user related data and security services data

- Strong Accountability

Any individual actor in an ATM network must hold full responsibility for its actions

- Activity Logging

The network elements must store information about security relevant events and operations

- Alarm Reporting

The network must be able to generate alarms on certain (configurable) security related events

- Audit

The network must support the capability to analyse and review the security relevant events in order to check them on violations of the security policy in force

- Security Recovery and Management of Security



Whenever an attempt to breach security occurs it must be possible to handle it in a controlled manner, that is with the minimum degradation of ATM network availability. Also, security management comprises all activities to establish, maintain and terminate security services.

The table below gives an overview of the principal functional security requirements and their mapping to the threats identified above.

Principal Functional Security Requirements	Generic Threats						
	Masquerade	Eavesdropping	Unauthorized Access	Corruption or Loss of Data	Repudiation	Forgery	Denial of Service
Verification of Identities	yes	--	yes	--	--	--	--
Controlled Access and Authorization	--	yes	yes	--	--	--	yes
Protection of Confidentiality	--	yes	yes	--	--	--	--
Protection of Data Integrity	--	--	--	yes	--	--	--
Strong Accountability	--	--	--	--	yes	yes	--
Activity Logging	yes	--	yes	--	yes	yes	yes
Alarm Reporting	yes	--	yes	yes	--	--	yes
Audit	yes	--	yes	--	yes	yes	yes
Security Recovery and Management of Security	yes	yes	yes	yes	yes	yes	yes

- The basic security services that will fulfill these requirements

Functional Security Requirement		Security Service
Verification of Identities		User Authentication Peer Entity Authentication Data Origin Authentication
Controlled Access and Authorization		Access Control
Protection of Confidentiality	of stored data	Access Control
	of data in transit	Confidentiality
Protection of Data Integrity	of stored data	Access Control
	of data in transit	Integrity
Strong Accountability		Non-repudiation
Activity Logging		Security Alarm, Audit Trail and Recovery
Alarm Reporting		Security Alarm, Audit Trail and Recovery
Audit		Security Alarm, Audit Trail and Recovery
Security Recovery and Management of Security		* see note

Note: This last requirement doesn't lead to a security service but determines the way the security services shall be specified.

- Plane Specific Interpretation of Functional Security Requirements

The basic functionality to be supported by each of the above mentioned services in general and specifically at the User Plane (and also, in part, at the Control Plane) is specified. Management Plane services are not elaborated upon. This includes the general methods used to implement the security services (namely, cryptographic algorithms with symmetric and single key), the options that shall be negotiable and the entities that may request the services.

- Support Services

This document enumerates the following Support Services required to assist plane specific security services:

- security message exchange protocols and basic negotiation
- security messaging in the control plane and user plane
- key exchange
- session key update
- certification

## ATM Security Specification, Version 1.0

This document provides the specification of some of the ATM security services of the User and Control Planes and Support Services. Specification of Management Plane security services is not provided, however, as management plane entities use user plane connections to perform their functions, the user plane security services will contribute for Management plane security.

The scope of the specification is restricted to ATM security, i.e.. mechanisms must be implemented in the ATM Layer or the AAL.

An implementation may claim compliance to this specification if it supports at least one security service (excluding support services), one of the required "profiles". For instance, one can say his implementation complies if he supports control plane authentication based on H-MD5 with key update through MD5.

Five services, identified in the Framework document examined above, are specified, together with their reference models:

- User Plane Security Services

These services apply on an end-to-end, per-VC basis (either Virtual Channel Connection or a Virtual Path Connection), not in physical links. Permanent and Switched VCs, point-to-point and point-to-multipoint are supported.

- Authentication (AUTH)

In the sense of "entity authentication" of the framework, certifies the identity of the calling and called parties at the beginning of the connection. It can be mutual or unilateral, and is based on the use of nonces and message digests.

- Confidentiality (CONF)

It's provided via encryption of data in transit. Encryption is applied to all or part of the payloads of the ATM cells - so it **takes place at the ATM layer**- and it is based on the use of symmetric key algorithms that are negotiated at call setup. Certain cells (namely, OAM and RM cells), though part of an encrypted user VC shall bypass the encryption process and be transmitted in clear.

- Integrity (INTEG)

Only available on virtual channels, not virtual paths. It shall be based on cryptographic checksums (or hashes, or message digests) appended to AAL3/4 and AAL5 "common part" data units. It is provided with two options: data integrity with replay/reordering protection, and data integrity without replay/reordering protection. The data integrity function takes place at the AAL.

- Access Control (ACC)

Performed during connection establishment. Access control decisions can be made as necessary at each ATM component in the connection path. Only one algorithm (Label-Based Access Control) is defined in the specification, but provision is made to support other mechanisms.

- Control Plane Security Services (CP-AUTH)

For the control plane, only a Data Origin Authentication and Integrity service is specified. Confidentiality and other security mechanisms are left out for later versions of the specification. The Data Origin Authentication and Integrity service is accomplished hop-by-hop between adjacent signalling elements, by applying the user plane data origin authentication and integrity services to the Signalling AAL connecting two signalling entities.

The provision of control plane data origin authentication and integrity, and the algorithms and parameters required to effect it are not dynamically negotiated, but configured by management. The option for replay and reordering protection of the data integrity service shall always be set.

Besides of these services, basic support services are identified and specified:

- security message exchange and negotiation of security options

Two mechanisms for the transport of security information are described in the specification: message exchange within **UNI 4.0 Signalling**(see [3]), and **inband** message exchange (that is, the message exchange takes place within the user VC, after it has been established by regular signalling). The choice transport mechanism influences the placement of the communicating Security Agents that will be involved in the establishment of the (user plane) security service. These Agents can be placed (in architecture terms) either as logical part of the control plane - in case Signalling is used, - or outside the control plane - when inband negotiation is used. Control plane security services shall always rely on UNI 4.0 negotiation.

Security message exchange can be accomplished through two-way and three-way protocols, but negotiation of parameters for security services is only possible through the a three-way exchange.

**Only two-way exchange protocols are specified with UNI 4.0 Signalling support.**

Each algorithm is associated with a "codepoint" used to identify it during negotiation of security options.

All mechanisms and protocols described use the **Security Services Information Element** which is

described in this document. This new IE conveys all security-related parameters and is transported in SETUP, and CONNECT signalling messages for point-to-point connection establishment; and ADD PARTY, and ADD PARTY ACKNOWLEDGE messages for point-to-multipoint connection establishment. For inband negotiation, three-way exchange protocols are specified, with the creation of a new message: CONNECT ACKNOWLEDGE. The Security Services IE is specified in detail in [3].

- key exchange

Is performed by using either symmetric key algorithms or asymmetric algorithms for encryption and authentication.

- key update

A Session Key Update Protocol is specified, which makes use of OAM cells within the user data stream to exchange the new session key with the remote party and to indicate when to start using the new key. The F4 and F5 OAM flows are used for VP and VC connections, respectively. The processes of updating and changing keys are independent in each direction.

- certification

ATM certification shall be based in X.509v1 certificates when certificate exchange occurs in the signalling channel. When certificate exchange is performed inband, X.509v1, X.509v2, and X.509v3 will be used.

The security message exchange protocol is used for exchanging public key certificates.

## UNI Signalling 4.0 Security Addendum

This document is still a Baseline Text from the Control Signalling Working Group. It specifies signalling procedures required to support the security services specified in [4], in addition to those described in UNI 4.0 Signalling [5].

It defines the Security Agent, an entity residing in the control plane responsible for processing security related information. As a result of security exchanges, the security agent can choose to accept or reject a call. The internal functionality of the security agent is not described in this document, it is covered by [4].

The most important modifications introduced to the signalling state machine has to do with the new Information Element (SSIE, described above) and the support for three-way exchanges in connection establishment. Basically, all signalling procedures are changed in order to include the Security Agent in the processing of the messages received, and the CONNECT ACKNOWLEDGE signalling message is added, and its processing defined.

## PHASE II

Since there are still no formal specifications available in this area there is no point in trying to determine if the equipment complies with them, however, current equipment provides some features that may be used in the enforcement of a security policy. Two mainstream enterprise switches were examined: the Fore ASX200BX, running ForeThought 5.1, and the Cisco LS1010, running IOS 11.3-0.8.TWA4.1.36. These features are briefly examined from the point of view of the generic security services they provide or support: Access Control, Authentication, Confidentiality, Integrity, and Logging (which may serve for audit or alarm reporting).

### Management Interface Security

The configuration tasks can, in both cases, be performed through a command-line interface or through SNMP. The command-line interface is accessible either by the console port of the switch or through the network, using TELNET.

As virtually all kinds of attack can benefit from the control of intermediate switches in the target ATM cloud, the security issues pertaining to the configuration interface of switching network elements are specially important.

The table below summarizes the security options of the two switches concerning the access to administrator-level privilege of the configuration interfaces.

		ASX200BX - FT5.1	LS1010 - IOS11.3
Access Control	SNMP	community-string, IP filters	SNMP specific access-lists
	Command-Line	User accounts with different privileges and IP filters	User accounts with different privileges and login specific access lists
Authentication	SNMP	none	SNMPv2
	Command-Line	optional password and SecurID	optional password and TACACS+
Logging	SNMP	none	syslog
	Command-Line	syslog, SNMP-TRAP	syslog, SNMP-TRAP
Confidentiality	SNMP	none	none
	Command-Line	password protection only, with SecurID	none

Integrity

none

none

In both switches, IP filters can be configured to limit the machines from which a user can log in or an SNMP access is allowed.

In the course of the testing of these switches some security flaws were detected:

- None of the switches provides any kind of encryption of the messages exchanged through the network while remotely configuring the machine.
- Most commands of the ASX200BX are available to "unprivileged" users

This kind of users, which implicitly should have almost read-only rights, can make a wide range of changes in the running configuration - including the establishment, sniffing, and deletion of PVCs

- The Fore's SNMP factory defaults are very loose. The read and write communities are set to publicly known strings and SNMP "Set Request" operations are enabled by default.

## ATM Call Filtering

The filtering of calls that are routed through a switch, specially when this switch is in the border between a private and a public network, is a very important (and effective) access control mechanism. If the filtering criteria were flexible enough (including bandwidth and QoS parameters, for instance), this facility could support the implementation of **Policy Based SVC Management** (see experiment 7), provided some authentication means are also available.

It is important to note that Call filtering without a reliable Data Origin Authentication Service is not enough to warrant access control. The case where a bogus NSAP is purposely registered by a calling station ILMI process is counteracted on the LS1010 through the use of special purpose access filters. Still, it may be possible to issue a SETUP request bearing a fake Caller NSAP address after registering a valid NSAP on the ingress switch...

Both switches support per-signalling interface call filtering, based on ordered lists of criteria and the direction of traversal of the interface by the call (incoming or outgoing). These criteria and additional capabilities are summarized below:

ASX200BX - FT5.1

LS1010 - IOS11.3

Criteria	Source NSAP address	yes, with nibble wildcards at the end of the address*	yes, with nibble and bit wildcard in the middle of the address
	Source NSAP mask	yes, with bit granularity	available through wildcards
	Destination NSAP address	yes, with nibble wildcards at the end of the address*	yes, with nibble and bit wildcard in the middle of the address
	Destination NSAP mask	yes, with bit granularity	available through wildcards
	Time of the day	no	yes, with minute granularity
	Call parameters	no	no
Logging	Violations	yes	no
	Statistics	yes	no

\* Fore claims the nibble wildcards work in intermediate positions of the address, but this was not the observed behaviour.

Besides of NSAP filtering, Cisco provides an implementation of the ITU-T Closed User Group signalling, specified in [2], though only a subset of the Interlock codes is supported.

As both switches support Lan Emulation and Classical IP as servers and clients, features that could address the security problems of these technologies were also examined.

Both switches support a validation mechanism in LANE, in which the NSAP of a client registering with the LES (when the LES is running on the switch, obviously) is sent to the LECS, which matches it to a mask of allowed addresses in the ELAN. If the NSAP does not match this mask, the LES refuses to accept the registration of the client in this ELAN.

For Classical IP none of the switches implements any kind of security measure.

The table below classifies these facilities according to the ATM Forum scheme:

	ASX200BX - FT5.1	LS1010 - IOS11.3
Access Control	LANE Membership restriction	LANE Membership restriction
		Closed User Groups



## References

1. ITU-T, Series I Recommendations, I.251.3 - I.251.7, 1995
2. ITU-T, Series I Recommendations, I.255.1 - Closed user group, 1992
3. ATM Forum Technical Committee, ATM Security Framework 1.0 (af-sec-0096.000 - Final Ballot in January 1998), December 1997
4. ATM Forum Technical Committee, UNI Signalling 4.0 Security Addendum (btd-sig-sec-01.02, baseline text), September 1997
5. ATM Forum Technical Committee, ATM Security Specification Version 1.0 (str-sec-01.02 - Straw Ballot in April 98), February 1998
6. ATM Forum Technical Committee, UNI Signalling 4.0, July 1996
7. Heinanen, J., Multiprotocol Encapsulation over ATM Adaptation Layer 5, RFC 1483, 1993
8. Laubach, M., Classical IP and ARP over ATM, RFC 1577, 1993
9. Perez, M., et al, ATM Signalling Support for IP over ATM, RFC 1755, 1995
10. Cole, R., et al, IP over ATM: A Framework Document, RFC 1932, 1996
11. Armitage, G., Support for Multicast over UNI 3.0 /3.1 based ATM Networks, RFC 2022, 1996
12. Luciani, J., et al, NBMA Next Hop Resolution Protocol (NHRP), INTERNET DRAFT (draft-ietf-rolc-nhrp-15.txt), 1998
13. Armitage, G., et al, Security issues for the ATMARP protocol, INTERNET DRAFT (draft-armitage-ion-sec-arp-00.txt), 1997
14. Armitage, G., Security issues for ION protocols, INTERNET DRAFT (draft-armitage-ion-security-01.txt), 1997

# Executive Summary

The TEN-34 project consists of two parts: A high-speed production network, and a research component that investigates new networking technologies and their applicability for the TEN-34 production network. The aim of this split has been to make standard high-speed networking capabilities available as soon as possible whilst allowing advances in technology to be incorporated at a later stage. The test programme was for organisational reasons split in two phases: WP 11 covered the first phase, and WP 14 the second. In deliverable D14.1 a test plan was defined for the second phase of the experiments, and this deliverable reports in detail on the results.

To avoid interference with the production TEN-34 network, all tests were carried out on a separate infrastructure. The pan-European JAMES ATM network was used for this purpose. The JAMES network provided ATM VP/VC facilities, which were used for the experiments. For most of them, an overlay network of VPs was created between the participating European research organisations to ease the management of connection changes for the different experiments.

The first phase of experiments (work package 11) examined various mostly ATM based technologies. The conclusion at the time was that whilst most of these new technologies such as SVCs or NHRP work in principle, the implementations were not stable enough yet for a production service. There were also a number of operational problems related with these new technologies. It was for example not possible to restrict the set-up of SVCs in a suitable fashion, to prevent un-authorized users from using network resources. Due to these problems no new features could be incorporated into the operational network. However, valuable experience was gained which will be useful for future deployment of new services.

The second phase of the experiments, the results of which are described in this deliverable, continued some of the experiments from phase one. It also investigated some new technologies, such as label based switching, which allows easy interaction of existing IP services with ATM technology. The overall results show that the stability of implementations has improved considerably, so that pilot services in some areas can be envisaged.

Some difficulties were found, however, in many cases those were not caused by the technology itself, but by other factors. Two key problems were identified in several experiments, which seem to be of a general nature:

First there is a clear lack of end-user applications that can in practice make use of new technologies. For example there are hardly any programs and network stacks making use of ATM natively, i.e., without IP. However, it is expected that the increasing availability of new technologies on the network side will boost new developments on the application side.

Secondly there is still a considerable lack of control and management functionality for new applications. This renders otherwise workable scenarios operationally impossible, since the use of expensive

resources such as international bandwidth has to be carefully policed and accounted for.

Overall the results present a concise overview of new networking technologies with applicability considerations and an outline of current weaknesses. Although no new features were implemented on the TEN-34 network due to the problems outlined, several technologies are promising and will be envisaged for future network developments.

---

[Back](#) to table of contents

# 1. Summary of Results per Experiment

1. **ATM Routing:** In this experiment a set of switches were configured first as a flat PNNI routing hierarchy, then as a multi-level peering hierarchy. Two types of ATM switches were used, and official NSAP addressing was applied. The results showed that ATM routing with PNNI could be used without problems over the test network. The network was overall stable and the set-up uncomplicated.
2. **Label based switching:** While this technology is being standardised, a proprietary implementation based on Cisco routers and switches was tested. Apart from some initial instabilities in the software, the experiment showed that this is a promising technology to provide a scalable architecture that combines the advantages of IP and ATM. More tests on high-load situations should be carried out, as well as more detailed comparisons with traditional IP networks.
3. **ATM resource reservation:** Due to limitations in hard- and software this test could not be carried out. There were no end-systems available that allow the specification of QoS parameters with defined traffic characteristics, and there were limitations for managing ATM connections inside shaped VP tunnels. This is for further study.
4. **IP resource reservation:** The tests of RSVP showed that this is now a stable technology for IP based reservation signalling, although policy based admission control is still required for full operational use. Interoperability between different vendors was working without problems. There are difficulties with RSVP on some lower layers such as shared media where reservations are by nature more difficult. More experiments are needed to test RSVP with various queuing techniques. The scalability of RSVP across service provider boundaries and for the Internet in general is very questionable due to technical as well as administrative reasons.
5. **ATM point-to-multipoint:** The basic functionality of unidirectional point-to- multipoint SVCs was demonstrated in a PIM sparse mode environment. More investigation is needed in the areas of bi-directional SVC trees, QoS signalling between end-systems using RSVP, and native ATM multicast using e.g. MARS. Other traffic classes but UBR could not be tested due to lack of available implementations and are left for further study.
6. **ATM signalling:** This experiment was continued from phase 1. Experiments with more recent software showed this time that SVCs can now be provided more reliably. The main problem for introduction in an operational network remains to be the difficulty to administratively police SVCs. There are some open questions with regard to the Signalling 4.0 standard, and SVCs of other traffic classes than UBR, both of which could not be tested as no implementations were available.
7. **ATM policy control and accounting:** This was a study on how to solve the administrative

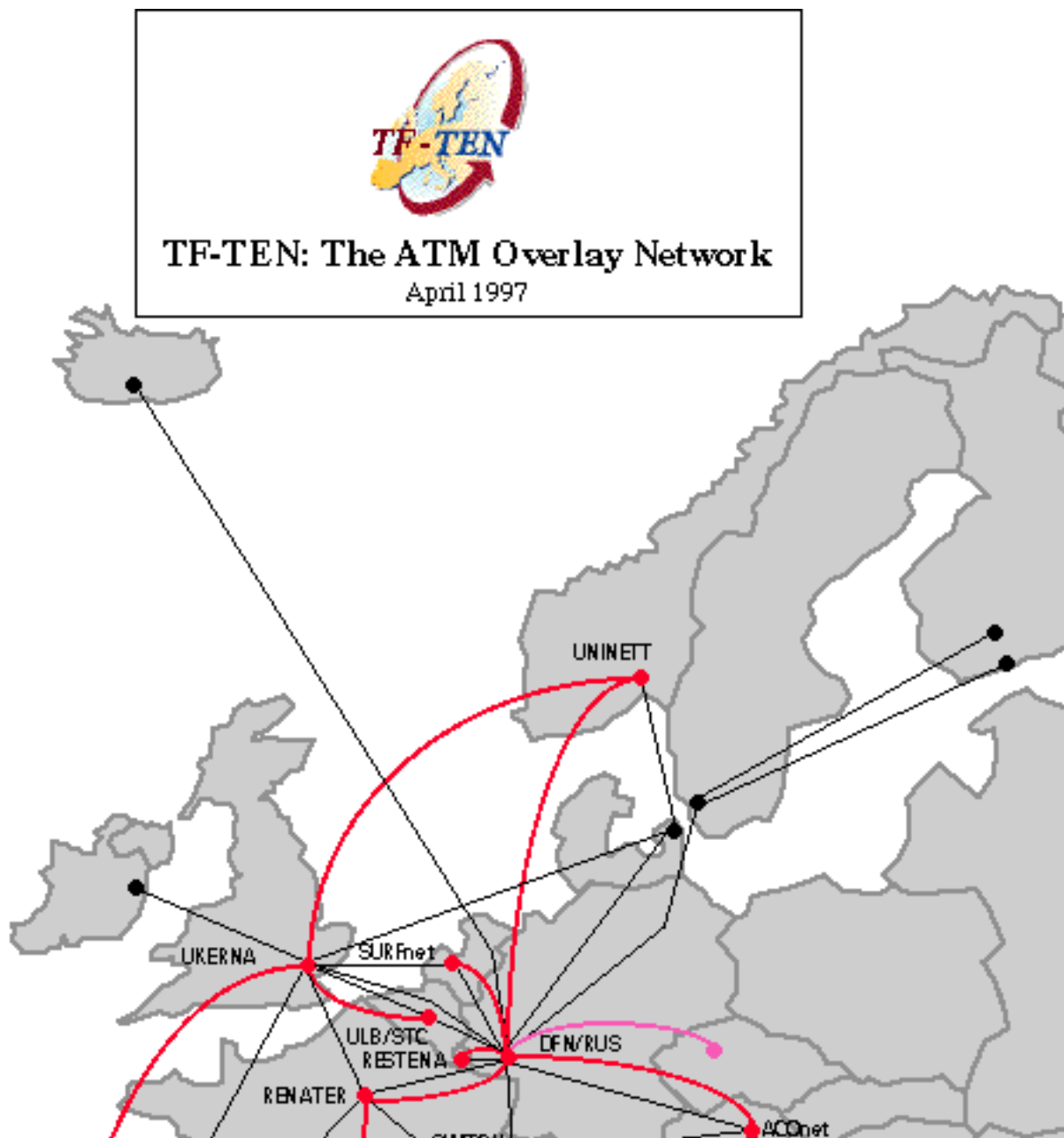
problems in the management of SVC. The investigation covered the set-up control, interference with existing SVCs, and accounting of used AVC resources. It gives recommendations on how to define policy decisions and examines the work of the IETF on RAP in the light of interworking with ATM.

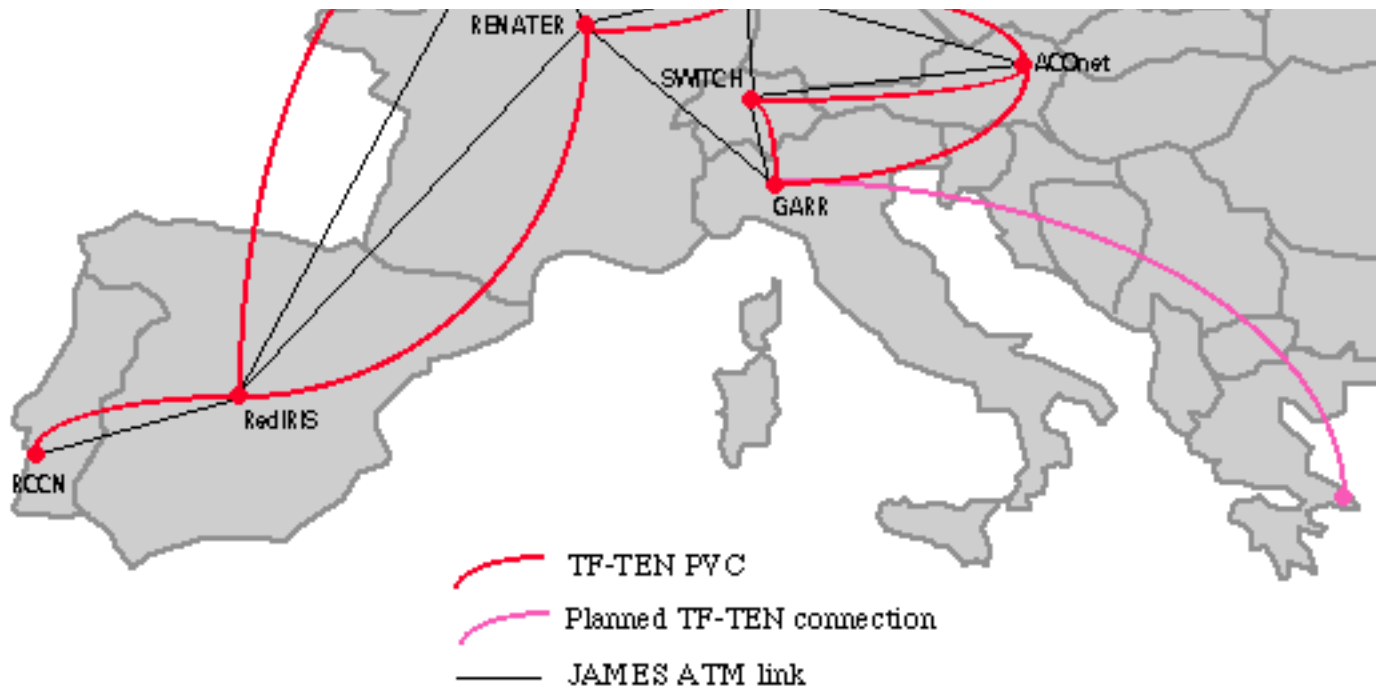
8. ATM traffic management: This experiment investigated various methods to provide the ABR traffic class, and tested a proprietary ABR implementation, since there were hardly any standardised ABR versions available at the time. An important result is that the resource management cells used by ABR are being handled correctly by the ATM switches. The tested ABR variant worked well, but is applicable only in a restricted environment.
9. ATM address resolution: Since ARP was tested in phase 1, this phase concentrated on NHRP on a statically routed network, and a network using OSPF. The tests revealed that NHRP is working well in both environments. There were some inconsistencies on the number of packets required to trigger a shortcut SVC. MPOA could not be tested, since no implementations were available.
10. ATM addressing: This activity concentrated on the use of public and private addressing schemes in an international environment. While public network operators are mostly preferring E.164 addresses, the European research networks are clearly favouring NSAP addressing. Potential problems of using two addressing structures were outlined, and the possibilities of address translation were examined, but practical tests are left for further study.
11. Native ATM performance: This experiment investigated native ATM applications and their performance. At the time there was no standardised ATM-API available, which meant that applications were specific to certain hardware and could not be easily ported. A native application (TCP-ONIP) was tested, which produced the expected performance.
12. ATM network management: SNMP based management of ATM devices showed that basic functions are working, but for example OAM MIBs are not standardised yet, and can only be used with proprietary MIBs. This activity provided an overview on currently supported MIBs. A web based network view was developed to monitor the status of the overlay network.
13. Security in ATM Networks: In this activity a threat model to advanced ATM services (with some form of signalling) was outlined, as well as mechanisms that are being developed currently to prevent potential attacks. It gives an overview about ongoing work in the security area in ATM. In addition the implementation of these mechanisms was examined on two types of switches.
14. Mesh Project: This project worked closely together with the native ATM experiment. It demonstrated video transmission over native high-speed ATM connections, without using IP.

## 2. Usage of the JAMES Network

The JAMES network provided the basic infrastructure over which the TEN-34 experiments described in this deliverable were carried out. The JAMES network ceased to exist on 31st of March 1998, at which time the experiments were concluded.

The set-up of connections over the JAMES network had to follow rigid administrative guidelines, with unspecified lead times, which made the planning of experiments difficult. For this reason the TERENA Task Force TEN (TF-TEN) requested a set of low-speed connections between the participating countries to build an overlay network that connected all participants to the experiments. This set-up allowed for short term changes to be made on the edge of the network, which was under the control of the TF-TEN. The overlay network was maintained by JAMES over the whole duration of the test programme, which made the planning of the experiments easier.





Technically the overlay network consisted of 2 Mbit/s (mostly 4750 cells/s) virtual paths of the CBR traffic class. When more bandwidth was required for certain experiments or demonstrations, additional higher speed VPs were requested from JAMES with a limited lifetime. Since the JAMES network did not fully support signalling, all experiments that required signalling were using VP tunnels. With this method all the signalling information is sent on the VP provided by JAMES instead of VP 0 in native ATM networks. This had the disadvantage of not being able to test ATM signalling natively across PNO networks, but it provided insight into the way the tunnelling of signalling information can be achieved.

---

[Back](#) to table of contents

### 3. Conclusions

The second phase of the experiments continued some of the experiments of the first phase, particularly those that had stability problems. Other experiments in this phase concerned latest developments in the IP and ATM world. In general the stability of implementations has improved considerably since the first test phase. For example the signalling tests showed significant instabilities during the first phase, to the extent that this technology could not be used in a multi-vendor WAN environment at the time. During this test phase the results were considerably better, although occasional problems could still be observed. This tendency could be observed with most experiments.

The interworking of IP and ATM is an essential requirement for many wide area networks, since IP technology is ubiquitously deployed at the edges of the Internet, where ATM does not have a sufficient coverage today. ATM on the other hand is being used frequently in the core of WANs and on campus networks. A number of technologies target to facilitate easy interworking of IP and ATM, to take advantage of both protocols. In this test phase emphasis was put on the interworking of IP and ATM, since it is becoming an essential part of network engineering.

In the core of such interworking techniques are RSVP and label based switching, which both aim at smooth interworking of IP over ATM or other connection oriented protocols. Both technologies are however not suitable for the global Internet, because the devices in the network are required to maintain a state for each data flow in the case of RSVP, which does not scale. Label based switching on the other hand relies on a flat hierarchy routing protocol, which in turn would not scale to the dimensions of the Internet. Both technologies however provide useful possibilities for providing differentiated services. RSVP implementations are quite reliable today, but there are still only few applications that can request QoS parameters. Label based switching is currently being standardised at the IETF. The experiments described here were carried out over proprietary implementations, which looked promising.

Another technology aiming at efficient IP transport over ATM is NHRP, which could be tested successfully on the overlay network. The implementations of NHRP did in general work stable, and produced mostly the expected results.

Some pure ATM technologies were also examined on their stability and correctness. PNNI routing for example worked well, although only with the latest versions of switch software. The resource reservation foreseen in PNNI could not yet be tested, because there were no implementations available yet. Point-to-multipoint SVCs could be tested and produced some good initial results, although more testing is required to fully understand the details of this technology. In addition, some experiments were done on ABR, with mostly proprietary implementations. The tests show that this is a difficult traffic class to realise in a WAN environment, due to the comparatively high delays. There are also only few implementations available yet. ABR in the WAN is therefore not yet fully understood and requires more investigation. Native ATM applications are scarce and testing was therefore very limited, although the few applications that could be tested did not show performance or stability problems.



Another set of experiments centered around the management of ATM networks. Whilst the principal methods are understood, the implementations to support advanced ATM management are not widely implemented yet, and require more work. Some alternative methods of monitoring the network were used, such as web based display of IP level reachability over the ATM network.

The implementations of switch software do not in all cases adapt to the highest security standards. During this test programme further security specifications for advanced techniques such as signalling were being developed and could therefore not yet been considered here.

In general, whilst the standard features of ATM are becoming stable with latest software releases, the new features are still not in a state to deploy them on a multi-vendor environment, where there are very high requirements in terms of availability or stability. This is a general problem that affected most experiments in some way. However, this is expected to be solved with future releases.

Another general problem in testing advanced ATM based services is the lack of end-user applications that can directly access the ATM stack of the computer, and request the service they require. This could be due to a lack of global roll-out of ATM to the desk-top, which would be required for general deployment of ATM applications.

Also the lack of operational control of the new ATM services is currently forcing businesses and universities to deploy standard technology. It is with almost no ATM feature possible to administratively restrict the usage of network resources. Also accounting of usage is still under development, and far from usable in a standard environment. This is a major blocking factor currently for the deployment of ATM, and whilst it is being addressed by vendors of equipment, it will take considerable time to be deployed on a sufficient scope.

In terms of traffic classes ATM equipment supports mostly only UBR in a sufficient manner for advanced features. This is enough for many applications, but since the driving force for ATM is the delivery of quality of service, this is disconcerting, because UBR does not offer QoS on its own. More development is needed also in this area.

In general the conclusion is that most of the advanced ATM features are very complex to implement, and take far more time than originally anticipated. This is to some extent due to the high variety of options in ATM, which put high demands on equipment and software. More work will need to be done in this area as more services will become available. The basic ATM transmission facilities such as CBR PVCs and PVPs work reliably and have provided an alternative to leased lines of delivering bandwidth.

In the meantime quality of service options are also being implemented on the IP protocol, making use of the precedence field present in each IP packet. This is a very simple mechanism which is unlikely to deliver the same hard guarantees as ATM. However, it might be sufficient for most applications. This is also for further study.

A concise summary of the results of this test phase will also be published in deliverable D14.3.

---

[Back](#) to table of contents

## 4. Detailed Test Descriptions

This section contains a detailed description of all experiments and activities carried out in the second phase of the TEN-34 test programme. The experiments carried out are:

- 4.1 [ATM Routing](#)
- 4.2 [Label based switching](#)
- 4.3 [ATM resource reservation](#)
- 4.4 [IP resource reservation](#)
- 4.5 [ATM point-to-multipoint](#)
- 4.6 [ATM signalling](#)
- 4.7 [ATM policy control and accounting](#)
- 4.8 [ATM traffic management](#)
- 4.9 [ATM address resolution](#)
- 4.10 [ATM addressing](#)
- 4.11 [Native ATM performance](#)
- 4.12 [ATM network management](#)
- 4.13 [Security in ATM Networks](#)
- 4.14 [The Mesh project](#)

---

[Back](#) to table of contents

## 4.1. ATM Routing

### Experiment Leader

Günther Schmittner, Johannes Kepler University, Linz

### Summary

PNNI is an ATM Forum specification for connecting either ATM nodes (switches) or ATM networks. *PNNI* stands for *Private Network-to-Network Interface* and has been approved in its current version PNNI 1.0 in March 1996. It consists of two categories of protocols, one for distributing topology and routing information between physical switches or groups of switches, based on well-known link-state routing techniques, the other for signalling point-to-point and point-to-multipoint connections across an ATM network, based on ATM Forum UNI signalling standards. PNNI supports source routing, crankback mechanisms and alternate routing of call setup requests in case of connection setup failure.

An experimental PNNI network has been set up on the TF-TEN overlay network (see section 2), which was also used as a platform for other experiments like ATM Signalling, ATM Point-to-Multipoint, ATM Address Resolution. PNNI proved to be a stable routing and signalling platform for a European ATM network. Most PNNI implementations on the market are quite new, some vendors provided early proprietary predecessors of ATM Forum PNNI 1.0. Today almost all of the vendor products (among them are Cisco, FORE, Bay, 3Com, Xylan, Olicom, IBM) only support a limited version of PNNI 1.0 in the sense of support for multiple hierarchies. There is only one implementation on the market (Cisco), we could get hands on for testing, which supports multi-level PNNI.

*Interim Inter-Switch Signalling Protocol (IISP)* (also known as PNNI Phase 0) implements an early ATM routing function using static routing tables in the ATM switches. IISP is aimed to interoperate with PNNI-based routing. Those switches in the network which did not support PNNI or were using incompatible versions or inappropriate network configurations, could be successfully integrated into the PNNI network.

The experiment was carried out in two phases, the first phase being to set up a stable PNNI based network running as a single peer group and to test dynamic routing and signalling functions. It is essential to design a proper ATM addressing structure before enabling PNNI in an ATM network. The different choices of ATM address types in the participating NRNs lead to a very uncommon PNNI configuration.

In the second phase multi-level PNNI switches were carefully introduced to test topology and routing information aggregation. A few incompatibilities with single-level implementations were detected, some of them could be fixed during the experiment. It should be noted that the used software was still beta-code and that the group had good support by the vendor for testing. Moreover, switches from a second vendor were integrated and interoperated successfully.

Integrated PNNI (I-PNNI) and public NNI (B-ICI) as complimentary and counterpart versions of PNNI could not be studied due to missing implementations and hardware.

In summary, PNNI could be used as a basis for a European ATM network infrastructure. Examples from some countries show that PNNI is already being used in campus environments or even at the NRN level for an ATM production network.

## Participants

JKU (Austria), SWITCH (Switzerland), RUS (Germany), RedIRIS (Spain), Renater (France), INFN (Italy), University of Twente (Netherlands), UNINETT (Norway), RCCN (Portugal), UKERNA (UK)

## Goals

- Study operation of ATM switches running PNNI in the WAN
- Prove interoperability between different PNNI implementations
- Verify interaction between PNNI-based and static ATM routing
- Gain experience in running a stable ATM routing infrastructure
- Prove the applicability of PNNI for a European ATM infrastructure
- Study the applicability of I-PNNI
- Compare functionality of PNNI with NNI

## Network infrastructure and description

PNNI (in the following short for ATM Forum PNNI 1.0) has been approved by the ATM Forum as an interface between ATM switches in the private network sector (an extended version, PNNI 2.0 is currently in work and is scheduled to be out 4/99).

The objectives in the design of PNNI can be summarized as follows:

- A functional interface between ATM switches such that full-function networks of arbitrary size and complexity may be constructed.
- Scalability
- Multi-Vendor
- Proprietary Subnetworks
- Open Implementation
- Dynamism
- Efficiency
- Usefulness

In order to accomplish those functions PNNI consists of following components:

- Topology and Routing
- Path Selection and Signalling
- Traffic Management
- Network Management

PNNI uses a hierarchical concept for its routing based on multiple (up to 104) topology levels. These levels are related to ATM addresses and make use of prefix-based routing. ATM switches running PNNI interact on each level independently among each other. The concept of *Logical Group Nodes (LGN)* simplify the exchange of routing information on the various layers.

When data connections are being set using PNNI signalling the ATM switches find and allocate a path through the network which satisfies the requested QoS characteristics. PNNI uses a source routing vector for this path which is

carried in the form of a *designated transit list (DTL)*. Mechanisms for alternate path selection are implemented using a *crankback protocol*.

Path selection and allocation is done using two different forms of *Connection Admission Control (CAC)*. These are called Generic Connection Admission Control (GCAC) and Actual Connection Admission Control (ACAC).

The experiment studied the operation of PNNI on various ATM switches in different NRNs over JAMES. As JAMES itself did not support PNNI or even PNNI interfaces for testing, we had to connect the private ATM switches over the JAMES infrastructure using VP bearer service. These VPs were of type CBR or VBR, in general at a bandwidth of 2 Mbps and are also called VP tunnels. The layout of the physical and logical infrastructure can be seen in section 2.

Of particular interest were the interoperability between different PNNI implementations, the interaction between PNNI-based and static ATM routing and the applicability of PNNI for a European ATM infrastructure.

As the efficient operation of PNNI is related to the allocation of ATM addresses, a careful planning of ATM address types and address allocation for and inside a NRN is needed.

In the first phase of TF-TEN experiments (see Deliverable D11.3) all private ATM switches had been connected using the predecessor of PNNI, Interim Inter-Switch Signalling Protocol (IISP). IISP or PNNI phase 0 is based on a modified UNI 3.0, whereby one of the switches reflects the network and the other the user side. Many results showed that due to different and limited implementations of IISP by various vendors, interoperability problems occur which in turn lead to signalling problems and failing call setup requests.

IISP uses static routing tables and is therefore cumbersome to maintain and to scale. However, ATM switches still support IISP and should interoperate with switches running PNNI. One can also run both on the same switch, some interfaces using PNNI others using IISP, supporting a smooth migration towards a pure PNNI network.

There are several implementations of PNNI on the market which should comply to ATM Forum PNNI 1.0 Specification completely or in part (e.g. single peer group support only). The TF-TEN group used switches from two different vendors, namely Cisco and FORE for testing. The actual products were:

- Cisco Lightstream 1010 (LS1010)
- FORE ASX-200BX
- FORE LE155

Today both vendors support a software version which supports a single hierarchy (one level) of PNNI speaking switches. The versions are

- IOS 11.1 or 11.2 for the LS1010
- ForeThought 5.1 for the FORE switches

Cisco made available to the group a beta version of IOS 11.3 code, which supports multiple peer-groups, that is more than one level of PNNI switches (the code has been officially released at the end of March).

## Network configurations

The PNNI experiment was carried out in two phases:

- Start with implementation of small PNNI islands and finally create a single level PNNI cloud among the participating countries
- Try to deploy a few multiple peer-group PNNI switches and link them to the big single level cloud

The TF-TEN overlay network provided a link bandwidth of 2 Mbps, which is by far sufficient to run PNNI.

As some other experiments were relying on a working SVC-based network, in particular the ATM Signalling, ATM Resource Reservation and ATM Addressing the network had to be stable and reliable.

The original time schedule for the experiment was a little too ambitious, as actual implementations by the various vendors were not available when we had planned to start.

In December 1997 we started preparation for a single level PNNI network. The only software to support it (at that time) was Cisco IOS 11.1 or 11.2. Johannes Kepler University had already some experience in running PNNI, because it had been used in the local and metropolitan area since January 1997 successfully in production on 11 Cisco LS1010 switches.

This release supports a single level (one hierarchy) only. As a consequence all private switches which wanted to participate had to be part of this single (lowest) level peer group. Following the PNNI specification a peer group consists of those switches which share a common prefix of bits in their ATM addresses.

Following ATM addresses are currently being used by the participating networks:

JKU (AT)	39.040F.54040101
ULB/STC (BE)	39.056F
SWITCH (CH)	39.756F.11111111700100011002
DFN/RUS (DE)	39.276F.3100011000000001
RedIRIS (ES)	39.724F.10010001000100010001
Renater (FR)	39.250F.0000002D
INFN (IT)	39.380F.00000000000000000000
CNAF (IT)	39.380F.10010001000000010000
RESTENA (LU)	39.442F
SURFnet (NL)	39.528F.1100
UNINETT (NO)	47.0023.01000005
RCCN (PT)	39.620F.00000000000000000000
UKERNA (UK)	39.826F.1107250010

Belgium and Luxemburg were not actively participating in the PNNI experiment because they did not have appropriate hardware. UKERNA tried to link their FORE switch at a later stage of the experiment, but did not succeed. SURFnet joined in the last few weeks successfully with a FORE LE155 switch.

What can be seen easily from the remaining partners is that all but Norway are using the DCC (first byte of hex 39)

type of ATM addresses, UNINETT uses the ICD (hex 47) format.

To fulfill the requirement of having a common bit prefix for all switches we had to choose a value of 1, because there is just the single leading zero bit common to all those address prefixes. As such the single peer group had to use a level of 1 for configuration of the PNNI processes.

Here is a sample of a PNNI configuration (single level) on a Cisco LS1010 switch:

```
hostname JKUASTE1
atm address 39.040f.5404.0101.0001.9999.0001.0060.83c4.a301.00
atm router pnni
  precedence pnni-remote-exterior 2
  background-routes-enable
  statistics call
  node 1 level 1 lowest
  redistribute atm-static
```

To define interfaces on a LS1010 switch for PNNI you have to specify:

```
interface ATM0/0/3
  description GDC APEX Test Switch
  no ip address
  no atm auto-configuration
  no atm address-registration
  no atm ilmi-enable
  atm pacing 2000 force
  atm iisp side user
  atm pvp 4
  atm pvp 5
  atm pvp 6
!
interface ATM0/0/3.4 point-to-point
  description Tunnel to Switch Zuerich
  no atm auto-configuration
  atm nni
!
interface ATM0/0/3.5 point-to-point
  description Tunnel to RUS Stuttgart
  no atm auto-configuration
  atm nni
  atm pvc 5 84 rx-cttr 101 tx-cttr 101 interface ATM0/0/3.4 4 84
  atm pvc 5 165 rx-cttr 101 tx-cttr 101 interface ATM0/0/3.4 4 165
!
interface ATM0/0/3.6 point-to-point
  description Tunnel to INFN Milano
  no atm auto-configuration
  atm nni
  atm pvc 6 164 rx-cttr 101 tx-cttr 101 interface ATM0/0/3.5 5 164
```



```
atm pvc 6 170 rx-cttr 101 tx-cttr 101 interface ATM0/0/3.5 5 170
```

The configuration above has been used on the Austrian switch at JKU for the connections towards CH, DE and IT. The statement `atm nni` specifies an interface type of PNNI. Please note that you can still define PVCs on interfaces being defined as PNNI.

## Results and findings

In December 1997 and January 1998 small islands of PNNI speaking switches were established, whereby existing IISP interfaces and static routing tables were replaced by PNNI and dynamic routing. It was found that static routing information is treated by default with a higher priority than dynamic routes learned via PNNI. In order to prefer dynamic routes, one has to explicitly lower the precedence of `pnni-remote-exterior` information (see corresponding `precedence pnni-remote-exterior 2` statement above).

Those switches which were not capable of running PNNI were linked using IISP as before. It is essential that static routing information to those IISP connected networks is only defined on the gateway PNNI switch and therefore introduced into dynamic routing in one place of the network. Obsolete static routes on all other PNNI switches have to be removed to provide stable dynamic routing.

ATM static routes like

```
atm route 39.040f.5404.0101.0001.9999.0002... ATM0/0/1
atm route 39.040f.5404.0101.0001.9999.0003... ATM0/0/2
```

can be automatically redistributed into PNNI using the `redistribute atm-static` command in the PNNI process.

At the end of January 1998 we had established a PNNI based network, covering major parts of Europe (see figure 1).

### TF-TEN PNNI Network (Phase 1)

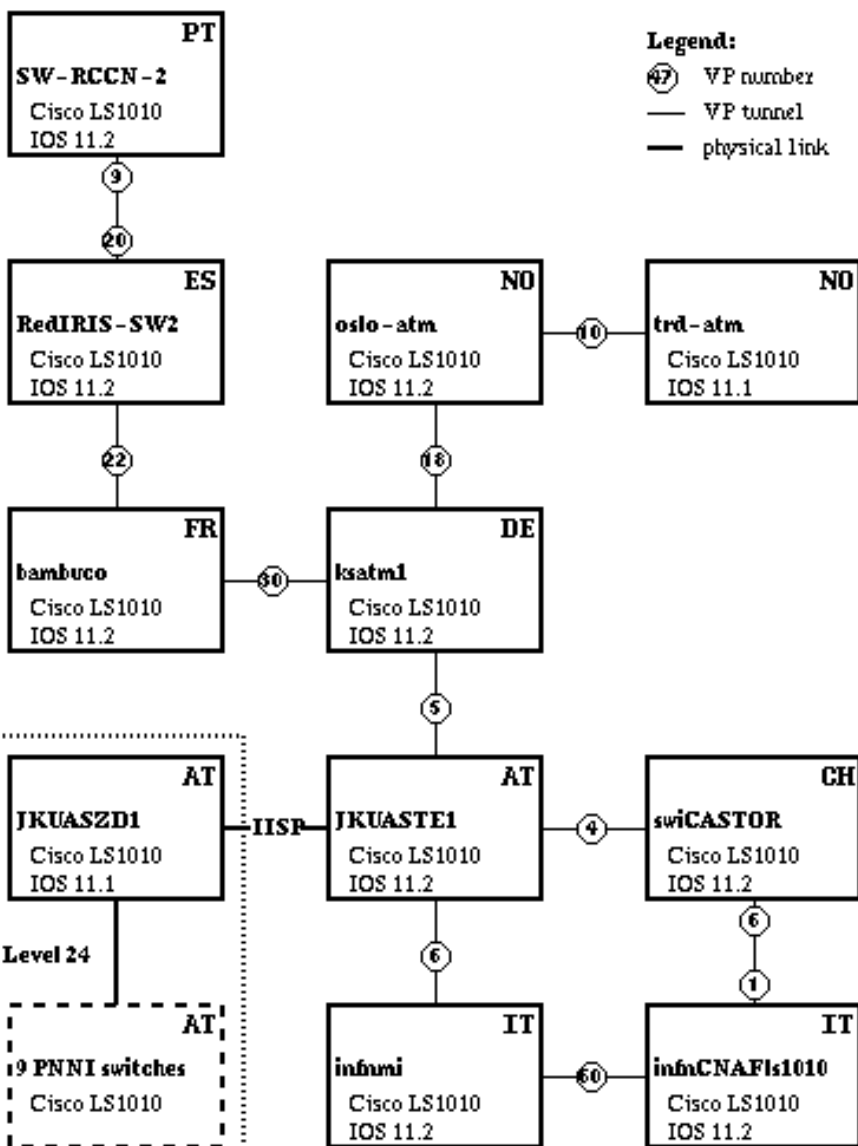


Fig: PNNI-1

Single peer group at level 1

38/01/00

Routing information is being distributed in a fully dynamic way and proved to be stable. This is a sample display of the ATM routing table as seen by node JKUASTE1:

```
JKUASTE1#sh atm route
```

Codes: P - installing Protocol (S - Static, P - PNNI, R - Routing control),  
 T - Type (I - Internal prefix, E - Exterior prefix, SE - Summary Exterior prefix, SI - Summary Internal prefix, ZE - Suppress Summary Exterior, ZI - Suppress Summary Internal)

P	T	Node/Port	St	Lev	Prefix
S	E	1 ATM0/0/0	UP	0	39.040f.5404.0101.0001/72
P	SI	1 0	UP	0	39.040f.5404.0101.0001.9999.0001/104

```

R I 1 ATM2/0/0 UP 0 39.040f.5404.0101.0001.9999.0001.0060.83c4.a301/152
R I 1 ATM0/1/0 UP 0 39.040f.5404.0101.0001.9999.0001.00e0.feb9.4c20/152
R I 1 ATM0/1/0 UP 0 39.040f.5404.0101.0001.9999.0001.00e0.feb9.4c21/152
R I 1 ATM0/1/0 UP 0 39.040f.5404.0101.0001.9999.0001.00e0.feb9.4c22/152
R I 1 ATM0/1/0 UP 0 39.040f.5404.0101.0001.9999.0001.00e0.feb9.4c23/152
R I 1 ATM2/0/0 UP 0 39.040f.5404.0101.0001.9999.0001.1111.1111.1111/152
R I 1 ATM2/0/0 UP 0 39.040f.5404.0101.0001.9999.0001.4000.0c/128
R I 1 ATM0/1/0 UP 0 39.040f.5404.0101.0001.9999.0001.9999.9999.9901/152
S E 1 ATM0/0/1 UP 0 39.040f.5404.0101.0001.9999.0002/104
S E 1 ATM0/0/2 UP 0 39.040f.5404.0101.0001.9999.0003/104
S E 1 ATM0/0/0 UP 0 39.040f.5404.0101.0002/72
S E 1 ATM0/0/0 UP 0 39.040f.5404.0101.0003/72
P E 10 0 DN 0 39.056f/24
P E 8 0 UP 0 39.250f.0000.002d.0014.0101.0001.0020.48/128
P I 6 0 UP 0 39.250f.0000.002d.0101.0101.0101/104
P I 3 0 UP 0 39.276f.3100.0110.0000.0001.0001/104
P E 3 0 UP 0 39.276f.3100.0110.0000.0001.0003/104
P E 3 0 UP 0 39.276f.3100.0110.0000.0001.0004/104
P I 8 0 UP 0 39.380f.0000.0000.0000.0000.0000/104
P E 8 0 UP 0 39.380f.0000.0000.0000.0000.0000.0000.f887.100e/152
P E 8 0 UP 0 39.380f.0000.0000.0000.0000.0000.0019.3246.0129/152
P E 8 0 UP 0 39.380f.0000.0000.0000.0000.0000.0019.3246.0130/152
P E 8 0 UP 0 39.380f.0000.0000.0000.0000.0000.0019.3246.0131/152
P E 8 0 UP 0 39.380f.0000.0000.0000.0000.0000.0019.3246.0132/152
P E 8 0 UP 0 39.380f.0000.0000.0000.0000.0000.0019.3246.0134/152
P E 8 0 UP 0 39.380f.0000.0000.0000.0000.0000.0020.4810.0958/152
P E 8 0 UP 0 39.380f.0000.0000.0000.0000.0000.0020.4815.15a9/152
P E 8 0 UP 0 39.380f.0000.0000.0000.0000.0000.0020.4816.02c1/152
P E 8 0 UP 0 39.380f.0000.0000.0000.0000.0000.0800.208b.2a00/152
P E 8 0 UP 0 39.380f.0000.0000.0000.0000.0000.8800.2b80.eda8/152
P I 11 0 UP 0 39.380f.1001.0001.0000.0001.0000/104
P I 9 0 UP 0 39.620f.0000.0000.0000.0000.0000/104
P E 9 0 UP 0 39.620f.0000.0000.0000.0000.0000/104
P I 7 0 UP 0 39.724f.1001.0001.0001.0001.0001/104
P I 2 0 UP 0 39.756f.1111.1111.7001.0001.1002/104
P I 10 0 DN 0 39.826f.1107.2500.1000.0000.0000/104
P E 10 0 DN 0 39.826f.1107.2500.1000.0000.0000.0020.4806.1ff1/152
P I 5 0 UP 1 47.0023.0100.0005.2000.0001.0101/104
P E 5 0 UP 1 47.0023.0100.0005.2000.0010/88
P I 4 0 UP 0 47.0023.0100.0005.4000.0001.0101/104
P E 4 0 UP 0 47.0023.0100.0005.4000.0001.0101.0020.4807.0979/152
P E 4 0 UP 0 47.0023.0100.0005.4000.0001.0101.0800.093d.063c/152

```

JKUASTE1#

PNNI is a link-state routing protocol and maintains up-to-date state information about the topology of the network. This is a typical topology display of the TF-TEN PNNI network:

JKUASTE1#sh atm pnni topology

Node 1 (name: JKUASTE1, type: ls1010, ios-version: 11.2)  
 Node Id: 1:160:39.040F54040101000199990001.006083C4A301.00  
 Service Classes Supported: UBR  
 Node Allows Transit Calls  
 Node has leadership priority 0

status	local port	remote port	neighbor
~~~~~	~~~~~	~~~~~	~~~~~
up	ATM0/0/3.4	ATM0/1/3.4	swiCASTOR
up	ATM0/0/3.5	ATM0/1/2.5	ksatml
up	ATM0/0/3.50	ATM0/1/2.50	ksatml
up	ATM0/0/3.52	ATM0/1/3.36	swiCASTOR
up	ATM0/0/3.6	ATM3/0/2.6	infnmi
up	ATM0/0/3.51	ATM0/0/3.51	infnCNAFls1010

Node 2 (name: swiCASTOR, type: ls1010, ios-version: 11.2)  
 Node Id: 1:160:39.756F11111111700100011002.00E014033F01.00  
 Service Classes Supported: UBR  
 Node Allows Transit Calls  
 Node has leadership priority 0

status	local port	remote port	neighbor
~~~~~	~~~~~	~~~~~	~~~~~
up	ATM0/1/3.4	ATM0/0/3.4	JKUASTE1
up	ATM0/1/3.36	ATM0/0/3.52	JKUASTE1
up	ATM0/1/3.37	ATM0/1/0.30	RedIRIS-SW2

Node 3 (name: ksatml, type: ls1010, ios-version: 11.2)  
 Node Id: 1:160:39.276F31000110000000010001.002048151140.00  
 Service Classes Supported: UBR  
 Node Allows Transit Calls  
 Node has leadership priority 0

status	local port	remote port	neighbor
~~~~~	~~~~~	~~~~~	~~~~~
up	ATM0/1/2.5	ATM0/0/3.5	JKUASTE1
up	ATM0/1/3.18	ATM1/1/1.18	oslo-atm
up	ATM0/1/2.50	ATM0/0/3.50	JKUASTE1
up	ATM0/1/3.30	ATM0/1/1.30	bambuco

Node 4 (name: oslo-atm, type: ls1010, ios-version: 11.2)  
 Node Id: 1:160:47.002301000005400000010101.00E014CB8A01.00  
 Service Classes Supported: UBR  
 Node Allows Transit Calls  
 Node has leadership priority 0

status	local port	remote port	neighbor
~~~~~	~~~~~	~~~~~	~~~~~

up	ATM1/1/1.18	ATM0/1/3.18	ksatml
up	ATM0/1/3.10	ATM0/1/3.10	trd-atm

Node 5 (name: trd-atm, type: ls1010, ios-version: 11.1)  
Node Id: 1:160:47.002301000005200000010101.00E014CB8701.00  
Service Classes Supported: UBR  
Node Allows Transit Calls  
Node has leadership priority 0

status	local port	remote port	neighbor
~~~~~	~~~~~	~~~~~	~~~~~
up	ATM0/1/3.10	ATM0/1/3.10	oslo-atm

Node 6 (name: bambuco, type: ls1010, ios-version: 11.2)  
Node Id: 1:160:39.250F0000002D0101010101.00E01E42EC01.00  
Service Classes Supported: UBR  
Node Allows Transit Calls  
Node has leadership priority 0

status	local port	remote port	neighbor
~~~~~	~~~~~	~~~~~	~~~~~
up	ATM0/1/1.30	ATM0/1/3.30	ksatml
up	ATM0/1/0.22	ATM0/1/0.22	RedIRIS-SW2

Node 7 (name: RedIRIS-SW2, type: ls1010, ios-version: 11.2)  
Node Id: 1:160:39.724F10010001000100010001.001011BBE301.00  
Service Classes Supported: UBR  
Node Allows Transit Calls  
Node has leadership priority 0

status	local port	remote port	neighbor
~~~~~	~~~~~	~~~~~	~~~~~
up	ATM0/1/0.20	ATM0/1/0.9	SW-RCCN-2
up	ATM0/1/0.30	ATM0/1/3.37	swiCASTOR
up	ATM0/1/0.22	ATM0/1/0.22	bambuco

Node 8 (name: infnmi, type: ls1010, ios-version: 11.2)  
Node Id: 1:160:39.380F00000000000000000000.00E014032400.00  
Service Classes Supported: UBR  
Node Allows Transit Calls  
Node has leadership priority 0

status	local port	remote port	neighbor
~~~~~	~~~~~	~~~~~	~~~~~
up	ATM3/0/2.6	ATM0/0/3.6	JKUASTE1

Node 9 (name: SW-RCCN-2, type: ls1010, ios-version: 11.2)  
Node Id: 1:160:39.620F00000000000000000000.001011BBD701.00  
Service Classes Supported: UBR

Node Allows Transit Calls

Node has leadership priority 0

status	local port	remote port	neighbor
~~~~~	~~~~~	~~~~~	~~~~~
up	ATM0/1/0.9	ATM0/1/0.20	RedIRIS-SW2

Node 10 (name: unknown, type: unknown, ios-version: unknown)

Node Id: 1:160:39.826F11072500100000000000.FF1A2E520001.00

Service Classes Supported:NONE

Node Allows Transit Calls

Node has leadership priority 0

status	local port	remote port	neighbor
~~~~~	~~~~~	~~~~~	~~~~~
up	ATM0/0/3.51	ATM0/0/3.51	JKUASTE1

Node 11 (name: infnCNAFls1010, type: ls1010, ios-version: 11.2)

Node Id: 1:160:39.380F10010001000000010000.001932460138.01

Service Classes Supported: UBR

Node Allows Transit Calls

Node has leadership priority 0

status	local port	remote port	neighbor
~~~~~	~~~~~	~~~~~	~~~~~
up	ATM0/0/3.51	ATM0/0/3.51	JKUASTE1

JKUASTE1#

PNNI provided a more stable signalling infrastructure than the IISP-based network in phase 1 of the experiments. This was particular important for the related experiments which had to rely on a stable SVC network.

We wanted to test interoperability between different vendor implementations but unfortunately the FORE software was not available at that time, so the network in phase 1 was only Cisco based. The single peer group software runs stable and releases 11.1 and 11.2 of the software can interact without problems.

According to the PNNI specifications switches on the lowest level communicate over physical links using the reserved VPI 0 and VCI 18. They exchange HELLO packets and topology state and routing information (PTSEs) over this logical link. In our case the physical links between the private switches were actually CBR VPs over the JAMES network (tunnels). Subsequently PNNI traffic was carried on VCI 18 within a particular VP, allocated by the PNOs.

Figure PNNI-1 shows the allocated VP numbers for the TF-TEN overlay network. In most cases the VP number was the same on both ends of the VP, in some rare cases it was different. PNNI is able to cope with different numbers on both ends of a tunnel, this is different to our experiences with IISP. In phase 1 of the TF-TEN experiments we had found that the Cisco implementation of IISP (on the LS100) required both numbers to be the same.

Another issue is the traffic generated by PNNI in the tunnel. Almost all of the public network operators enforce traffic policing at the entry point to their public ATM network. Thus we had to make sure, that we did not violate the traffic contract for the 2 Mbps connections, in particular not to send too high bursts of cells into JAMES. We implemented traffic shaping on the physical interface towards the PNO to limit the traffic sent by PNNI and by other applications of

course.

The `atm pacing 2000 force` command on the LS1010 makes sure that outgoing traffic is shaped at 2 Mbps. Unfortunately this command can only be applied to physical interfaces. This is a problem for all those sites which run more than one tunnel, because tunnels on the LS1010 are implemented as logical interfaces on a physical interface (the one towards the carrier). Therefore traffic shaping using `atm pacing 2000` effectively limits the sum of all tunnels to 2 Mbps. More information regarding this topic can be found in the *ATM Resource Reservation* experiment.

In February we started to test multi-level (hierarchical) PNNI. As said earlier it is supported on the LS1010 with IOS 11.3. Cisco provided beta code for 11.3 for this experiment. IOS 11.3 requires at least 32 MB DRAM on the switch, which was also provided as a loan from Cisco Europe.

At the same time FORE released ForeThought 5.1 software for their ForeRunner switch family, which supports single level PNNI 1.0 as well as a multi-level version of FORE's proprietary ForeThought PNNI (FT-PNNI). FORE switches at JKU, RCCN and University of Twente were successfully integrated into the level 1 PNNI cloud during February/March. Another FORE switch at UKERNA could not be placed online due to unresolved configuration problems.

A switch running multiple levels on PNNI actually becomes member of multiple peer groups. Each instance of a PNNI process within the same switch runs at a distinct level. As at the lowest level all switches at a specific level share the same bit prefix. One of the main ideas of PNNI is the concept of logical group nodes (LGN). A group of switches forming a peer group at a certain level can be represented at the next higher level as a single virtual node or logical group node. The actual switch which acts as a representation for a group of switches at the lower level is called peer group leader (PGL). The PGL aggregates topology and routing information in the upward direction (to the next higher level) and distributes it downwards to the switches it represents (to the next lower level). It serves as a simplification of a part of the network, not only for routing but also for call setup and resource allocation. In that way PNNI networks can scale up to a size of thousands of switches.

The main goal of the second phase of our PNNI experiment was to establish a two-level hierarchy at certain points in the network, study interaction between the levels and verify that route summarization (aggregation) and signalling works correctly. To achieve that, we had to run 11.3 beta code on at least two LS1010 switches, define a two-level hierarchical PNNI node on them (ideally using different lower levels on both) and see how it works.

JKU started to run a two-level PNNI node at levels 24 and 1. The highest level had to be 1 for obvious reasons. Johannes Kepler University and two other institutions in the metropolitan area of Linz run altogether 10 PNNI switches (9 Cisco LS1010s and a FORE ASX-200BX) in production, sharing a 24 bit address prefix. Therefore the lowest level of the test switch at JKU (JKUASTE1) was defined at level 24. This is the relevant part of the configuration:

```
atm address 39.040f.5404.0101.0001.9999.0001.0060.83c4.a301.00
atm router pnni
precedence pnni-remote-exterior 2
background-routes-enable
statistics call
node 1 level 24 lowest
  parent 2
  redistribute atm-static
  election leadership-priority 10
node 2 level 1
```

```
election leadership-priority 20
name Austria
```

IOS 11.3 currently supports a maximum of 8 levels (out of a theoretical maximum of 104), the number after node is an identifier of a PNNI process at a certain level. The level statement defines the bit level at which the node operates and the parent statement is the reference number of the next higher level process. The name command can be used to identify the LGN.

It is essential to specify an election leadership-priority value, otherwise the node will not be eligible to become peer group leader.

This is a sample display of a running multi-level PNNI node:

```
JKUASTE1#sh atm pnni hierarchy
  Locally configured parent nodes:
  Node          Parent
  Index  Level  Index  Local-node Status  Node Name
  ~~~~~  ~~~~~  ~~~~~  ~~~~~~
  1      24     2      Enabled/ Running  JKUASTE1
  2      1     N/A     Enabled/ Running  Austria
JKUASTE1#sh atm pnni local-node
```

PNNI node 1 is enabled and running

```
Node name: JKUASTE1
System address      39.040F54040101000199990001.006083C4A301.01
Node ID            24:160:39.040F54040101000199990001.006083C4A301.00
Peer group ID      24:39.040F.0000.0000.0000.0000.0000
Level 24, Priority 10 60, No. of interfaces 5, No. of neighbors 1
Parent Node Index: 2
Node Allows Transit Calls
Node Representation: simple
```

```
Hello interval 15 sec, inactivity factor 5,
Hello hold-down 10 tenths of sec
Ack-delay 10 tenths of sec, retransmit interval 5 sec,
Resource poll interval 5 sec
SVCC integrity times: calling 35 sec, called 50 sec,
Horizontal Link inactivity time 120 sec,
PTSE refresh interval 1800 sec, lifetime factor 200 percent,
Min PTSE interval 10 tenths of sec
Auto summarization: on, Supported PNNI versions: newest 1, oldest 1
Default administrative weight mode: uniform
Max admin weight percentage: -1
Next resource poll in 3 seconds
Max PTSEs requested per PTSE request packet: 32
Redistributing static routes: Yes
```

PNNI node 2 is enabled and running



```

Node name: Austria
System address      39.040F54040101000199990001.006083C4A301.02
Node ID            1:24:39.040F00000000000000000000.006083C4A301.00
Peer group ID      1:00.0000.0000.0000.0000.0000.0000
Level 1, Priority 10 0, No. of interfaces 0, No. of neighbors 0
Parent Node Index: NONE
Node Allows Transit Calls
Node Representation: simple

```

```

Hello interval 15 sec, inactivity factor 5,
Hello hold-down 10 tenths of sec
Ack-delay 10 tenths of sec, retransmit interval 5 sec,
Resource poll interval 5 sec
SVCC integrity times: calling 35 sec, called 50 sec,
Horizontal Link inactivity time 120 sec,
PTSE refresh interval 1800 sec, lifetime factor 200 percent,
Min PTSE interval 10 tenths of sec
Auto summarization: on, Supported PNNI versions: newest 1, oldest 1
Default administrative weight mode: uniform
Max admin weight percentage: -1
Max PTSEs requested per PTSE request packet: 32
Redistributing static routes: No

```

```
JKUASTE1#
```

Soon after establishment of the first multi-level PNNI node at JKU it turned out that communication with the neighbor switches had been lost. After some investigation and with the help of Cisco we found that a bug in IOS 11.2(10) (and earlier releases) code prevented the PNNI links from coming up. The bug only prevented interaction between a multi-level 11.3 and a 11.2 system, single-level 11.3 and 11.2 switches could work together. This was the reason why the link between JKU (JKUASTE1) and INFN Milano (infnmi) in figure PNNI-2 below ceased to work during this phase of the experiment.

Another bug we discovered was that connections from 11.3 connected end systems towards 11.2 connected ones succeeded, but in the other direction failed. A special version of 11.2 code fixed that problem (fix should be incorporated in the 11.2(12) release).

Here is an example output of JKUASTE1 and it's connections to the neighbors:

```
JKUASTE1#sh atm pnni interface
```

```
PNNI Interface(s) for local-node 1 (level=24):
```

Local Port	Type	RCC	Hello	St	Deriv	Agg	Remote Port	Rem Node(No./Name)
ATM0/0/0	Phy	UP	2way_in	0			ATM0/1/0	16 JKUASZD1
ATM0/0/3.4	VP	UP	comm_out	0			ATM0/0/2.4	20 swiCASTOR
ATM0/0/3.5	VP	UP	comm_out	0			ATM0/1/3.5	30 DE-ksatm1
ATM0/0/3.6	VP	UP	1way_out	0			ATM3/0/2.6	23 infnmi

```
PNNI Interface(s) for local-node 2 (level=1):
```

```

Local Port      Type  RCC HrzLn St Deriv Agg  Remote Port  Rem Node(No./Name)
~~~~~
2256000        Hrzn UP   2way  0      1440000      20 swiCASTOR
29CB000        Hrzn UP   2way  0      1440000      30 DE-ksatml
    
```

JKUASTE1#sh atm pnni neighbor

Neighbors For Node (Index 1, Level 24)

```

Neighbor Name: JKUASZD1, Node number: 16
Neighbor Node Id: 24:160:39.040F54040101000100010001.006083C4A201.00
Neighboring Peer State: Full
Link Selection Set To: minimize blocking of future calls
  Port          Remote Port Id  Hello state
  ATM0/0/0      ATM0/1/0        2way_in      (Flooding Port)
    
```

Neighbors For Node (Index 2, Level 1)

```

Neighbor Name: DE-ksatml, Node number: 30
Neighbor Node Id: 1:160:39.276F31000110000000010001.111111111111.00
Neighboring Peer State: Full
Link Selection Set To: minimize blocking of future calls
  Port          Remote Port Id  Hello state
  FFFFFFFF      FFFFFFFF        2way_in      (Flooding Port)
    
```

```

Neighbor Name: swiCASTOR, Node number: 20
Neighbor Node Id: 1:160:39.756F1111111700100011002.00E014033F01.00
Neighboring Peer State: Full
Link Selection Set To: minimize blocking of future calls
  Port          Remote Port Id  Hello state
  FFFFFFFF      FFFFFFFF        2way_in      (Flooding Port)
    
```

JKUASTE1#

A list of connected nodes and an excerpt of the topology database is shown below:

JKUASTE1#sh atm pnni identifiers

Node	Node Id	Name
1	24:160:39.040F54040101000199990001.006083C4A301.00	JKUASTE1
2	1:24:39.040F00000000000000000000.006083C4A301.00	Austria
9	24:160:39.040F54040200000000000004.00603E5B4C01.00	ls1010-log
10	24:160:39.040F54040200000000000003.00603E5B5201.00	ls1010-eg
11	24:160:39.040F54040200000000000002.0002DC602901.00	ls1010-wan
12	24:160:39.040F54040200000000000001.00603E5B4D01.00	ls1010-lan
13	24:160:39.040F54040101000100040001.006083C4A301.00	JKUASP11
14	24:160:39.040F54040101000100020001.001011B87B01.00	JKUASTT1
15	24:160:39.040F54040101000100010003.FF1A2DE80001.00	
16	24:160:39.040F54040101000100010001.006083C4A201.00	JKUASZD1
17	24:160:39.040F54010101000100010001.006083C4B101.00	HFG-LS1010
18	24:160:39.040F54040200000000000006.006083C5C200.00	ls1010-fl
19	24:160:39.040F54040200000000000005.006083C44E01.00	ls1010-2og

```

20      1:160:39.756F11111111700100011002.00E014033F01.00 swiCASTOR
21      1:160:47.00918100000000603E5ABF01.00603E5ABF01.00 tagswitchINFN
22      1:160:39.380F10010001000000010000.001932460138.01 infnCNAFls1010
23      1:160:39.380F00000000000000000000.00E014032400.00 infnmi
24      1:160:39.620F00000000000000000000.FF1A37140001.00
26      1:160:39.620F00000000000000000000.001011BBD701.00 SW-RCCN-2
27      1:160:39.724F10010001000100010001.001011BBE301.00 ATM-SW-MAD2
28      1:160:47.002301000005400000010101.00E014CB8A01.00 oslo-atm
29      1:160:47.002301000005200000010101.00E014CB8701.00 trd-atm
30      1:160:39.276F31000110000000010001.111111111111.00 DE-ksatml
31      1:160:39.250F0000002D010101010101.00E01E42EC01.00 bambuco

```

JKUASTE1#sh atm pnni topology

```

Node 1 (name: JKUASTE1, type: ls1010, ios-version: 11.3)
Node ID.: 24:160:39.040F54040101000199990001.006083C4A301.00
Node AESA: 39.040F54040101000199990001.006083C4A301.01
Link Service Classes Advertised: CBR VBR-RT VBR-NRT ABR UBR
Leadership Priority: 60, Claims PGL: Yes, Transit Calls: Allowed
Ancestor: No, Nodal Representation: Simple

```

status	link-type	local port	remote port	neighbor
up	uplink	ATM0/0/3.4	FFFFFFFF	swiCASTOR
up	hrz	ATM0/0/0	ATM0/1/0	JKUASZD1
up	uplink	ATM0/0/3.5	FFFFFFFF	DE-ksatml

```

Node 2 (name: Austria, type: ls1010, ios-version: 11.3)
Node ID.: 1:24:39.040F00000000000000000000.006083C4A301.00
Node AESA: 39.040F54040101000199990001.006083C4A301.02
Link Service Classes Advertised: UBR
Leadership Priority: 0, Claims PGL: No, Transit Calls: Allowed
Ancestor: Yes, Nodal Representation: Simple

```

status	link-type	local port	remote port	neighbor
up	hrz	2256000	1440000	swiCASTOR
up	hrz	29CB000	1440000	DE-ksatml

```

Node 16 (name: JKUASZD1, type: ls1010, ios-version: 11.2)
Node ID.: 24:160:39.040F54040101000100010001.006083C4A201.00
Node AESA: 39.040F54040101000100010001.006083C4A201.00
Link Service Classes Advertised: CBR VBR-RT VBR-NRT ABR UBR
Leadership Priority: 0, Claims PGL: No, Transit Calls: Allowed
Ancestor: No, Nodal Representation: Simple

```

status	link-type	local port	remote port	neighbor
up	hrz	ATM0/1/0	ATM0/0/0	JKUASTE1
up	hrz	ATM0/0/1	ATM0/0/0	JKUASP11

```

up      hrz      ATM0/0/2      ATM0/0/0      JKUASTT1
up      hrz      ATM0/0/0.21   ATM0/1/3.21   HFG-LS1010
up      hrz      ATM0/0/3      10000000      15
up      hrz      ATM0/0/0.22   ATM0/0/0.22   ls1010-wan

```

```

Node 20 (name: swiCASTOR, type: ls1010, ios-version: 11.3)
Node ID.: 1:160:39.756F11111111700100011002.00E014033F01.00
Node AESA: 39.756F11111111700100011002.00E014033F01.01
Link Service Classes Advertised: UBR
Leadership Priority: 0, Claims PGL: No, Transit Calls: Allowed
Ancestor: No, Nodal Representation: Simple

```

```

status  link-type  local port      remote port      neighbor
~~~~~  ~~~~~~
up      dnlink     FFFFFFFF        ATM0/0/3.4      JKUASTE1
up      hrz        ATM0/1/0.6      ATM0/1/0.1      tagswitchINFN
up      hrz        1440000         2256000         Austria

```

```

Node 30 (name: DE-ksatml, type: ls1010, ios-version: 11.3)
Node ID.: 1:160:39.276F31000110000000010001.111111111111.00
Node AESA: 39.276F31000110000000010001.111111111111.01
Link Service Classes Advertised: UBR
Leadership Priority: 0, Claims PGL: No, Transit Calls: Allowed
Ancestor: No, Nodal Representation: Simple

```

```

status  link-type  local port      remote port      neighbor
~~~~~  ~~~~~~
up      dnlink     FFFFFFFF        ATM0/0/3.5      JKUASTE1
up      hrz        ATM0/1/3.30     ATM0/1/1.30     bambuco
up      hrz        ATM0/1/3.18     ATM1/1/1.18     oslo-atm
up      hrz        1440000         29CB000         Austria

```

```
JKUASTE1#
```

SWITCH established a second multi-level PNNI node, whereby one of the lower level switches was actually a test switch in Bologna using an ATM address of SWITCH. This node (swiCASTOR) ran at levels 1 and 96, the LGN name at level 1 being ZH-Bologna.

```

atm router pnni
no e164-aesa
background-routes-enable
statistics call
node 1 level 96 lowest
parent 2
redistribute atm-static
election leadership-priority 10
name Zurich
node 2 level 1
election leadership-priority 20
name ZH-Bologna

```

```
swiCASTOR#sh atm pnni topology
```

```
Node 1 (name: Zurich, type: ls1010, ios-version: 11.3)
Node ID.: 96:160:39.756F11111111700100011002.00E014033F01.00
Node AESA: 39.756F11111111700100011002.00E014033F01.01
Link Service Classes Advertised: UBR
Leadership Priority: 60, Claims PGL: Yes, Transit Calls: Allowed
Ancestor: No, Nodal Representation: Simple
```

status	link-type	local port	remote port	neighbor
up	hrz	ATM0/1/0.6	ATM0/1/0.1	tagswitchINFN
up	uplink	ATM0/0/2.4	FFFFFFFF	Austria

```
Node 2 (name: ZH-Bologna, type: ls1010, ios-version: 11.3)
Node ID.: 1:96:39.756F11111111700100011000.00E014033F01.00
Node AESA: 39.756F11111111700100011002.00E014033F01.02
Link Service Classes Advertised: UBR
Leadership Priority: 0, Claims PGL: No, Transit Calls: Allowed
Ancestor: Yes, Nodal Representation: Simple
```

status	link-type	local port	remote port	neighbor
up	hrz	2440000	29F7000	Austria

[...]

Germany and Portugal ran single-level 11.3 nodes.

Between Austria and Switzerland we had a two-level hierarchical PNNI link. The LGN in Zurich (ZH-Bologna) effectively hides all switches at the lower level as shown below:

```
JKUASTE1#sh atm pnni identifiers
```

Node	Node Id	Name
1	24:160:39.040F54040101000199990001.006083C4A301.00	JKUASTE1
2	1:24:39.040F00000000000000000000.006083C4A301.00	Austria
9	24:160:39.040F54040200000000000004.00603E5B4C01.00	ls1010-log
10	24:160:39.040F54040200000000000003.00603E5B5201.00	ls1010-eg
11	24:160:39.040F54040200000000000002.0002DC602901.00	ls1010-wan
12	24:160:39.040F54040200000000000001.00603E5B4D01.00	ls1010-lan
13	24:160:39.040F54040101000100040001.006083C4A301.00	JKUASPI1
14	24:160:39.040F54040101000100020001.001011B87B01.00	JKUASTT1
15	24:160:39.040F54040101000100010003.FF1A2DE80001.00	
16	24:160:39.040F54040101000100010001.006083C4A201.00	JKUASZD1
17	24:160:39.040F54010101000100010001.006083C4B101.00	HFG-LS1010
18	24:160:39.040F54040200000000000006.006083C5C200.00	ls1010-fl
19	1:160:47.002301000005400000010101.00E014CB8A01.00	oslo-atm
<b>20</b>	<b>1:96:39.756F11111111700100011000.00E014033F01.00</b>	<b>ZH-Bologna</b>
24	1:160:39.620F00000000000000000000.001011BBD701.00	SW-RCCN-2

```

26      1:160:39.276F31000110000000010001.111111111111.00 DE-ksatm1
27      1:160:39.250F0000002D010101010101.00E01E42EC01.00 bambuco
28      1:160:39.620F00000000000000000000.FF1A37140001.00
29      1:160:39.528F11002000002000012005.FF1A4EB50002.00
30      1:160:39.724F10010001000100010001.001011BBE301.00 ATM-SW-MAD2
31      1:160:47.002301000005200000010101.00E014CB8701.00 trd-atm

```

JKUASTE1#

The link-type field in the topology display above shows values of hrz, uplink and dnlink, which correspond to the types of links between hierarchical nodes in the PNNI specification. Horizontal links are established between neighboring nodes on the same level, up and down links between neighbor nodes from different peer groups and the PGLs of the other peer group. All logical links except those on the lowest level over physical paths are actually SVCs. These connections are dynamically built as PNNI creates the hierarchy, forms peer groups and elects peer group leaders.

The corresponding routing table at JKUASTE1 is shown below:

JKUASTE1#sh atm route

Codes: P - installing Protocol (S - Static, P - PNNI, R - Routing control),  
T - Type (I - Internal prefix, E - Exterior prefix, SE -  
Summary Exterior prefix, SI - Summary Internal prefix,  
ZE - Suppress Summary Exterior, ZI - Suppress Summary Internal)

P	T	Node/Port	St	Lev	Prefix
~	~	~	~	~	~
P	I	17 0	UP	0	39.040f.5401.0101.0001.0001.0001/104
P	I	2 0	UP	0	39.040f.5401.0101.0001.0001.0001/104
P	E	16 0	UP	0	39.040f.5404.0101.0001.0000/88
P	E	2 0	UP	0	39.040f.5404.0101.0001.0000/88
P	I	16 0	UP	0	39.040f.5404.0101.0001.0001.0001/104
P	I	2 0	UP	0	39.040f.5404.0101.0001.0001.0001/104
P	I	15 0	UP	0	39.040f.5404.0101.0001.0001.0003/104
P	I	2 0	UP	0	39.040f.5404.0101.0001.0001.0003/104
P	E	16 0	UP	0	39.040f.5404.0101.0001.0001.0004/104
P	E	2 0	UP	0	39.040f.5404.0101.0001.0001.0004/104
P	I	14 0	UP	0	39.040f.5404.0101.0001.0002.0001/104
P	I	2 0	UP	0	39.040f.5404.0101.0001.0002.0001/104
P	E	14 0	UP	0	39.040f.5404.0101.0001.0002.0002/104
P	E	2 0	UP	0	39.040f.5404.0101.0001.0002.0002/104
P	I	13 0	UP	0	39.040f.5404.0101.0001.0004.0001/104
P	I	2 0	UP	0	39.040f.5404.0101.0001.0004.0001/104
P	SI	1 0	UP	0	39.040f.5404.0101.0001.9999.0001/104
P	I	2 0	UP	0	39.040f.5404.0101.0001.9999.0001/104
R	I	ATM0/1/3	UP	0	39.040f.5404.0101.0001.9999.0001.0020.4810.108a/152
R	I	ATM0/1/1	UP	0	39.040f.5404.0101.0001.9999.0001.0020.ea00.0b22/152
R	I	ATM2/0/0	UP	0	39.040f.5404.0101.0001.9999.0001.0060.83c4.a301/152
R	I	ATM0/1/0	UP	0	39.040f.5404.0101.0001.9999.0001.00e0.feb9.4c20/152

R	I	1	ATM0/1/0	UP	0	39.040f.5404.0101.0001.9999.0001.00e0.feb9.4c23/152
R	I	1	ATM2/0/0	UP	0	39.040f.5404.0101.0001.9999.0001.1111.1111.1111/152
R	I	1	ATM0/1/0	UP	0	39.040f.5404.0101.0001.9999.0001.1234.5678.90ab/152
R	I	1	ATM2/0/0	UP	0	39.040f.5404.0101.0001.9999.0001.4000.0c/128
R	I	1	ATM0/1/0	UP	0	39.040f.5404.0101.0001.9999.0001.9999.9999.9901/152
S	E	1	ATM0/0/1	DN	0	39.040f.5404.0101.0001.9999.0002/104
S	E	1	ATM0/0/2	UP	0	39.040f.5404.0101.0001.9999.0003/104
P	E	2	0	UP	0	39.040f.5404.0101.0001.9999.0003/104
P	I	15	0	UP	0	39.040f.5404.0101.0001.9999.0004/104
P	I	2	0	UP	0	39.040f.5404.0101.0001.9999.0004/104
P	E	16	0	UP	0	39.040f.5404.0101.0002/72
P	E	2	0	UP	0	39.040f.5404.0101.0002/72
P	E	16	0	UP	0	39.040f.5404.0101.0003/72
P	E	2	0	UP	0	39.040f.5404.0101.0003/72
P	I	12	0	UP	0	39.040f.5404.0200.0000.0000.0001/104
P	I	2	0	UP	0	39.040f.5404.0200.0000.0000.0001/104
P	I	11	0	UP	0	39.040f.5404.0200.0000.0000.0002.0002.dc60.2901/152
P	I	2	0	UP	0	39.040f.5404.0200.0000.0000.0002.0002.dc60.2901/152
P	I	11	0	UP	0	39.040f.5404.0200.0000.0000.0002.0060.3e26.9220/152
P	I	2	0	UP	0	39.040f.5404.0200.0000.0000.0002.0060.3e26.9220/152
P	I	11	0	UP	0	39.040f.5404.0200.0000.0000.0002.0800.093d.123f/152
P	I	2	0	UP	0	39.040f.5404.0200.0000.0000.0002.0800.093d.123f/152
P	I	11	0	UP	0	39.040f.5404.0200.0000.0000.0002.4000.0c/128
P	I	2	0	UP	0	39.040f.5404.0200.0000.0000.0002.4000.0c/128
P	I	10	0	UP	24	39.040f.5404.0200.0000.0000.0003/104
P	I	9	0	UP	24	39.040f.5404.0200.0000.0000.0004/104
P	I	18	0	UP	0	39.040f.5404.0200.0000.0000.0006/104
P	I	2	0	UP	0	39.040f.5404.0200.0000.0000.0006/104
P	I	27	0	UP	0	39.250f.0000.002d.0101.0101.0101.0020.4804.f850/152
P	I	27	0	UP	0	39.250f.0000.002d.0101.0101.0101.0020.4808.1ac3/152
P	I	27	0	UP	0	39.250f.0000.002d.0101.0101.0101.00e0.1e42.ec01/152
P	I	27	0	UP	0	39.250f.0000.002d.0101.0101.0101.1921.6801.3003/152
P	I	27	0	UP	0	39.250f.0000.002d.0101.0101.0101.1921.6801.4001/152
P	I	27	0	UP	0	39.250f.0000.002d.0101.0101.0101.1921.6801.6001/152
P	I	27	0	UP	0	39.250f.0000.002d.0101.0101.0101.1931.4618.5002/152
P	I	27	0	UP	0	39.250f.0000.002d.0101.0101.0101.1931.9615.2236/152
P	I	27	0	UP	0	39.250f.0000.002d.0101.0101.0101.1932.4600.0115/152
P	I	27	0	UP	0	39.250f.0000.002d.0101.0101.0101.1932.4600.0116/152
P	I	27	0	UP	0	39.250f.0000.002d.0101.0101.0101.4000.0c/128
P	I	26	0	UP	0	39.276f.3100.0110.0000.0001.0001.0020.4806.093f/152
P	I	26	0	UP	0	39.276f.3100.0110.0000.0001.0001.0020.480e.0243/152
P	I	26	0	UP	0	39.276f.3100.0110.0000.0001.0001.0020.ea00.059f/152
P	I	26	0	UP	0	39.276f.3100.0110.0000.0001.0001.1111.1111.1111/152
P	I	26	0	UP	0	39.276f.3100.0110.0000.0001.0001.4000.0c/128
P	E	26	0	UP	0	39.276f.3100.0110.0000.0001.0003/104
P	E	26	0	UP	0	39.276f.3100.0110.0000.0001.0004/104
P	I	29	0	UP	0	39.528f.1100.2000.0020.0001.2005/104
P	I	28	0	UP	0	39.620f.0000.0000.0000.0000.0000/104
P	I	24	0	UP	0	39.620f.0000.0000.0000.0000.0000.0010.11bb.d701/152

```

P I 24 0 UP 0 39.620f.0000.0000.0000.0000.0000.4000.0c/128
P I 28 0 UP 0 39.724f.0000.0000.0000.0000.0000/104
P I 30 0 UP 0 39.724f.1001.0001.0001.0001.0001/104
P I 28 0 UP 0 39.724f.1001.0001.0001.0001.0001/104
P E 30 0 UP 0 39.724f.1001.0001.0001.0001.0001.1034.1034.1034/152
P E 30 0 UP 0 39.724f.1001.0001.0001.0001.0001.1932.4600.0101/152
P I 20 0 UP 0 39.756f.1111.1111.7001.0001.1002/104
P I 20 0 UP 0 39.756f.1111.1111.7001.0001.1006/104
P I 28 0 UP 0 39.826f.1107.2500.10/64
P E 11 0 UP 0 47.0005.80ff.e100.0000/72
P E 2 0 UP 0 47.0005.80ff.e100.0000/72
P I 28 0 UP 0 47.0023.0100.0005/56
P I 31 0 UP 1 47.0023.0100.0005.2000.0001.0101/104
P E 31 0 UP 1 47.0023.0100.0005.2000.0010/88
P I 19 0 UP 0 47.0023.0100.0005.4000.0001.0101/104
P I 28 0 UP 0 47.0079.0000.0000.0000.0000.0000.00a0.3e00.0001/152

```

JKUASTE1#

It can be seen that two routing entries for 39.756f.1111.1111.7001.0001.1002/104 and 39.756f.1111.1111.7001.0001.1006/104 exist, both via node 20 (ZH-Bologna) although they should have been summarized at level 96 into one entry by the PGL.

This was actually another bug we discovered. It was quickly fixed by Cisco and after that the routing table looked correct:

JKUASTE1#sh atm route

Codes: P - installing Protocol (S - Static, P - PNNI, R - Routing control),  
T - Type (I - Internal prefix, E - Exterior prefix, SE -  
Summary Exterior prefix, SI - Summary Internal prefix,  
ZE - Suppress Summary Exterior, ZI - Suppress Summary Internal)

```

P T Node/Port St Lev Prefix
~ ~ ~~~~~~ ~ ~ ~~~~~~
P SI 2 0 UP 0 39.040f/24

```

[...deleted...]

```

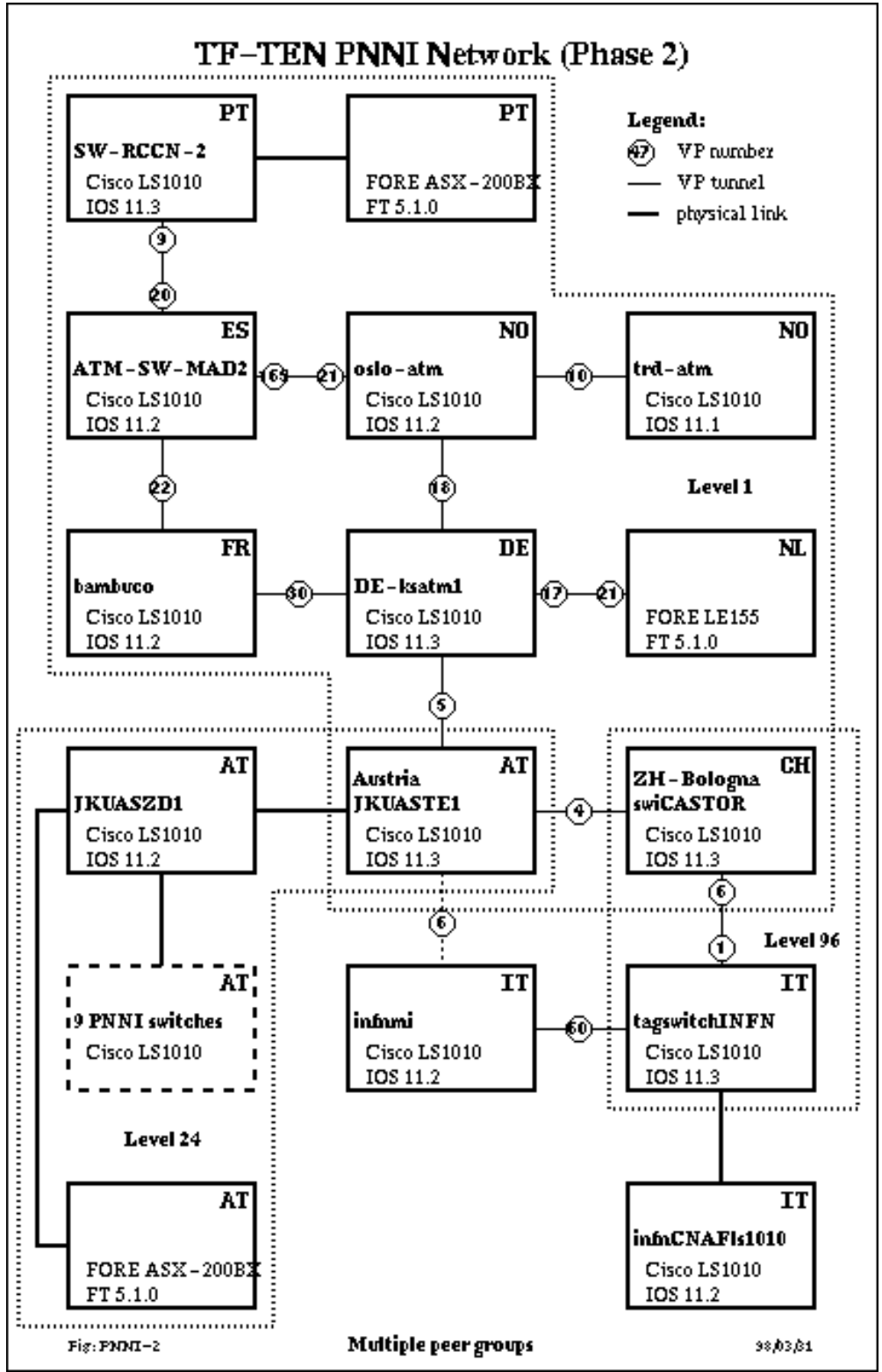
P I 20 0 UP 0 39.276f.3100.0110.0000.0001.0001.0020.ea00.059f/152
P I 20 0 UP 0 39.276f.3100.0110.0000.0001.0001.1111.1111.1111/152
P E 20 0 UP 0 39.276f.3100.0110.0000.0001.0001.2222.2222.2201/152
P I 20 0 UP 0 39.276f.3100.0110.0000.0001.0001.4000.0c/128
P I 26 0 UP 0 39.276f.3100.0110.0000.0001.0003/104
P E 25 0 UP 0 39.528f.1100.2000/56
P I 25 0 UP 0 39.528f.1100.2000.0020.0001.2005/104
P I 21 0 UP 0 39.756f.1111.1111.7001.0001.10/96
P E 11 0 UP 0 47.0005.80ff.e100.0000/72
P E 2 0 UP 0 47.0005.80ff.e100.0000/72

```



JKUASTE1#

Below is a picture of the final PNNI network at the end of March. It consisted of 23 PNNI nodes (20 Cisco LS1010s and 3 FORE switches) and two hierarchy levels. *Austria* and *ZH-Bologna* represented as LGN a couple of other switches running at levels 24 and 96.



Because of time constraints we could not extensively test the network. In the last few days we discovered an unusual

behaviour during connection setup, whereby certain calls only succeeded after some trials.

Another problem which occurred lately were asymmetric call setup failures between some end systems. Calls in one direction always failed, while in the other direction they always succeeded. There was no time to investigate further the reason for that problem, especially as it seemed not easily reproducible.

In general it was shown that PNNI can provide a stable platform for ATM routing and signalling on the European scale. However, careful planning of address allocation is a must to make best use of the scaling facilities PNNI provides. It was proven that PNNI implementations of different vendors interoperate. Different path selection criteria as supported by the implementations were not tested.

Related information can be found in the *ATM Resource Reservation* experiment.

## 4.2. Label-Based Switching: Architecture and Performance in an ATM Wide Area Network

### Experiment Leaders

Jean-Marc Uzé, RENATER, Paris, France  
Tiziana Ferrari, INFN/CNAF, Bologna, Italy

### Table of contents

[Introduction](#)

[Participants](#)

[Protocol overview](#)

[Phase 1: preliminary tests in laboratory](#)

[Phase 2: WAN tunnelling test](#)

[Phase 3: Experiment on the TF-TEN/JAMES infrastructure](#)

1. [Testbed](#)
2. [IP architecture for scalability](#)
3. [Performance](#)
4. [Traffic engineering](#)

[Conclusions and Future Work](#)

[Bibliography](#)

### Introduction

For Internet service providers (ISPs) the advantages of being able to scale the Internet across long distance WANs using ATM include:

- increased forwarding and routing capacity for public services
- improved traffic management between core Internet sites
- easier support for new Internet services
- more flexibility to adapt to evolving needs of Internet users

Network technologies need to be enhanced to support new applications and to cope with the increasing number of users. Increasing the availability of network resources is just not enough to achieve this goal: Scalable network architecture, increased packet forwarding capabilities, a wider range of services are all additional requirements. MultiProtocol Label Switching (MPLS) is one of the new networking techniques under study in the scientific community which aims at combining the flexibility of the IP protocol routing scheme with the speed typical of the cell switching technology.

We present the design of a wide area network based on MPLS and its implementation in an ATM testbed. In particular the paper aims at analysing the functionality and the performance of MPLS and at discussing the its applicability for the implementation of an ATM-based backbone.

The goal was first to get familiar with this technology, but also to study how this technology can be used to implement a scalable and simple wide area network interconnecting many exterior networks.

The MPLS test programme [1] was developed by the task force TF-TEN [2] of TERENA on the ATM European backbone *JAMES* [3] connecting European National Research Networks and co-founded by the European Community. The testbed connected several European countries: Austria (ACONET), France (RENATER), Germany (DFN), Italy (GARR), Spain (REDIRIS) and Switzerland (SWITCH).

The whole experiment has been performed in three phases to get gradually familiar with the technology and validate some configurations before setting up the whole network.

Section 2 provides an overview of MPLS functionalities and of the implementation used for the tests.

Section 3 details the first phase done in laboratory while Section 4 describes the second phase performed on a PNO infrastructure.

Section 5 describes the main experiment performed by all TF-TEN participants of the experiment, and details the configuration of the testbed (par. 5.1), the IP architecture set-up for scalability (par. 5.2), the performance tests (par. 5.3), and traffic engineering (par. 5.4).

## Participants

ACONET (Austria), DFN (Germany), GARR/INFN (Italy), RedIRIS (Spain), RENATER (France) and SWITCH (Switzerland).

## Protocol overview

The label-based switching [4][5] is a technique which integrates layer 2 switching and layer 3 routing. It is designed for high speed networks to efficiently make use of ATM performances and of IP routing scalability and flexibility. The idea is to set up ATM VCs for traffic that does not need to be routed at each hop. Implementations can be based on flow detection (e.g. IP switching) or routing topology (e.g. Tag switching). This technique can interoperate with several protocols but we will study only the applicability for an IP service on ATM.

For the tests we used a proprietary implementation of MPLS developed by Cisco and called tag switching. It was launched in September 1996 and the first beta versions were available during summer 1997.

Tag switching [6] is a technique that assigns a label or "tag" to packets towards their final destination. In a conventional router network, the packets are processed by each router on the path. With tag switching, the ingress router assigns a label to each destination network. The packets are then switched through this network by these tags towards the egress router. A tag switching network consists of a core of tag switches with tag edge routers on the periphery. Tag edge routers and tag switches use standard routing protocols (BGP, OSPF...) to build routing tables. Then, a tag (a couple VPi/VCi in ATM networks) is assigned to each route in the tag network. The tag information is distributed by the Tag Distribution Protocol (TDP) to all tag switches and tag edge routers. TDP uses a control pvc automatically set-up between adjacent tag devices during the initialization phase. This protocol will set-up all VCs in the ATM backbone to forward each IP flows.

## Phase 1: Preliminary tests in laboratory

This experiment was performed in France on September 1997, the 3<sup>rd</sup> and 4<sup>th</sup>.

### 1. Goals of Phase 1

The main goals of the phase one were to get familiar with the tag switching configuration on the LightStream 1010 and the Cisco 7500, to understand the features set and to prepare phase 2.

### 2. Configuration

For the test performed in laboratory we used the following configuration:

- 1 x LightStream 1010 32 Mbytes of memory, Feature Card 1, 4 ports ATM 155 Mb/s MMF

- 2 x Cisco 7505, 1 x EIP-6, 1 x AIP 155 Mb/s
- Software versions were beta and pre-beta images

The names for the switch and the routers are Tag-1010, Tag-Top and Tag-Bottom respectively.

### 3. Network Configuration

The network configuration is shown in Figure 1.

Each Cisco 7505 acting as Tag Edge router was connected to a LightStream 1010 acting as Tag Switch router. IP routing protocol was OSPF on all systems.

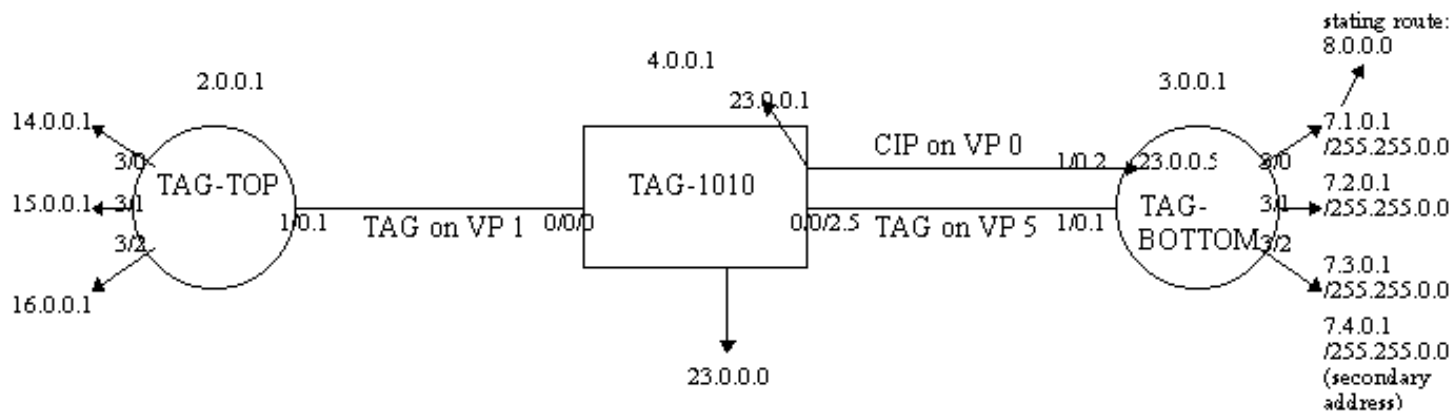


Figure 1: Phase 1 network infrastructure

- Cisco 7505 Tag-Top is configured with 3 IP networks (14.0.0.0, 15.0.0.0, 16.0.0.0) on Ethernet interfaces; the ATM interface uses interface IP unnumbered Loopback0 (network 2.0.0.0). Note: the minimum AIP hardware revision for Cisco 7500 is 1.3.

- Cisco 7505 Tag-Bottom is configured with 1 IP network (7.0.0.0) subnetted on Ethernet interfaces (e3/0 is 7.1.0.0, e3/1 is 7.2.0.0, e3/2 is 7.3.0.0 as well as secondary 7.4.0.0), the ATM interface uses interface IP unnumbered Loopback0 (network 3.0.0.0) for tag switching but sub-interface ATM1/0.2 is configured as ATM ARP Server for Classical IP (RFC 1577) on network (23.0.0.0) to demonstrate ATM native and tag switching on a single ATM interface as well as ATM signalling on LightStream 1010 ATM Switch. A static route for 8.0.0.0 is set on interface ethernet3/0.

- LightStream 1010 Tag-1010 is configured with 1 IP network (22.0.0.0) on Ethernet; ATM interfaces use interface IP unnumbered Loopback0 (network 4.0.0.0) for tag switching but ATM0/0/0 is set for direct connection with Tag-Top while ATM0/0/2 is configured with one sub-interface for tag switching using VPI 5 with Tag-Bottom. An other sub-interface is used by Classical IP from Tag-Bottom ATM2/0/0.2 is configured as ATM ARP Client for Classical IP (RFC 1577) on network (23.0.0.0) to demonstrate ATM native and tag switching on a single ATM Switch.

### 4. Tests performed

After configuring tag switching, we tested the different commands to check if it was up and running from both a tag switching and an IP connectivity view.

Cisco IOS CLI commands for tag switching are :

Tag-Bottom#sho tag ?

atm-tdp	ATM Tagging Protocol information
forwarding-table	Show the Tag Forwarding Information Base (TFIB)
interfaces	Show per-interface tag switching
tdp	Tag Distribution Protocol information
tsp-tunnels	Tag Switched Path tunnel status and configuration

Tag-Bottom#sho tag tdp ?

bindings	Show the TDP Tag Information Base (TIB)
discovery	Display sources for locally generated TDP Discovery Hello PDUs
neighbor	Display TDP neighbor information

These commands permit to check tag switching interfaces and neighbours, display tag switching Virtual Circuit (TVC) and monitor IP routes and IP route mapping

Also, to demonstrate that native ATM applications can run at same time as tag switching, Classical IP was set on one Cisco 7505 running an ARP Server and LightStream 1010. The constraint is that Tag must have a dedicated VP. In our test, Tag used VP 5 and Classical IP VP 0.

## 5. Results

Connectivity was observed by pinging the interfaces of the routers and switch.

We observed also some instabilities with these first test releases:

- TVCs remaining in bind-wait state on a router. This means that TVCs were not established between the router and the destinations concerned.
- sometimes we needed to reboot the switch and/or routers because check information was not available through the user interface commands.

The experiment was tested with OSPF in the the Tag network core (BGP was not available in that release).

We observed that there was no aggregation of IP routes. This means that for each destination announced by an egress TSR (Tag Switching Router), a TVC is established from each ingress TSR (scalability problem).

The ATM-specific VC merge feature was only available with a future release of Feature card (FC-PFQ). In any case, these card would not solve the aggregation problem on IP level.

This test was very interesting as it showed us that the use of this technology in a wide area backbone requires a specific IP architecture.

## Phase 2: WAN tunnelling test

### 1. Goals of Phase 2

The second phase was done in France on October 1997, the 1<sup>st</sup> and 2<sup>nd</sup>. We interconnected two sites, University of Jussieu (Paris) and Idris laboratory (Orsay) through a public ATM 10 Mbps CBR VP provided by the public network operator.

The main goal was to test the connectivity through an ATM public operator offering a VP service, in order to prepare the full deployment on JAMES (Phase 3). Actually in order to run tag switching on an ATM infrastructure just offering CBR PVP service, tag switching VCs (including the control VC) have to be tunnelled into the PVPs offered by the Public Network Operators, so tag switching is completely transparent for the ATM equipment on the public network side.

### 2. Configuration

- Paris-Jussieu:

- 1 x Cisco 7505 with EIP-6 and AIP 155Mb/s MMF
- 1 x LightStream 1010 with 4 ports OC-3

- Orsay-Idris:

- 1 x Cisco 7505 with EIP-6 and AIP 155Mb/s MMF
- 1 x LightStream 1010 with 4 ports OC-3

- Software versions:

Cisco 7505: tag switching EFT version

LightStream 1010: pre-beta version

The names for the switches and the routers are ls1010-orsay, 7505-orsay, ls1010-jussieu, and 7505-jussieu.

### 3. Network configuration

The network configuration is shown in figure 2.

Each Cisco 7505 acting as Tag Edge router was connected to a LightStream 1010 acting as Tag Switch router. The IP routing protocol was OSPF on all systems.

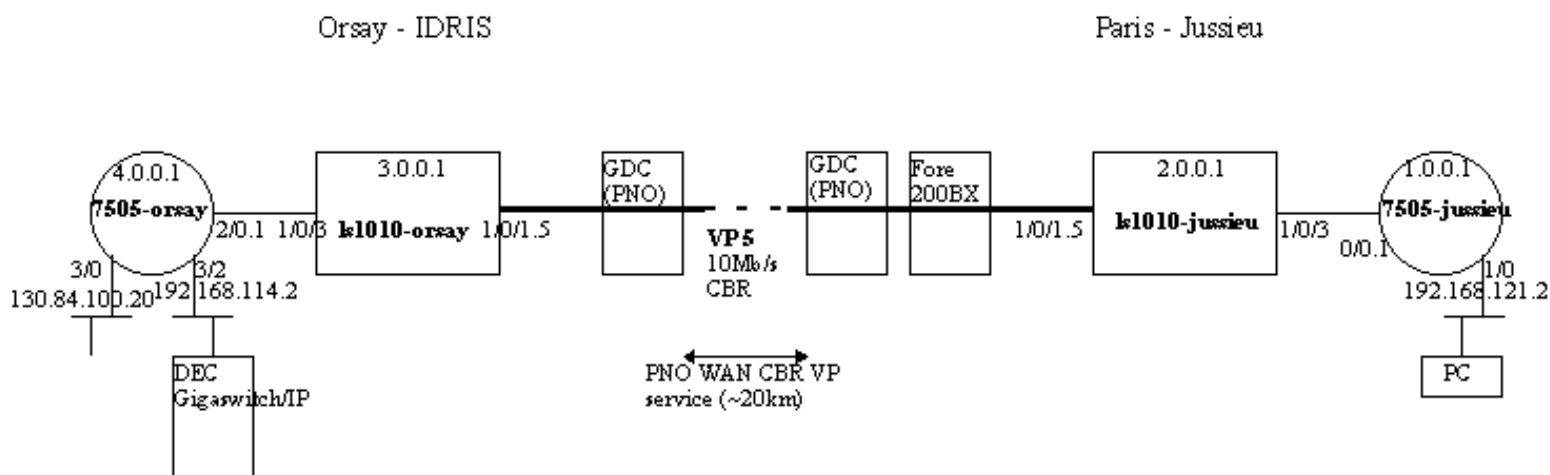


Figure 2: Phase 2 network infrastructure

Orsay-Idris Ethernet segments were connected to Cisco 7505 Ethernet E3/0 and E3/2; OSPF adjacency on E3/2 was created with DEC Gigaswitch/IP (Ipsilon) installed at Idris. ATM 155 Mb/s AIP connected the router to the LightStream 1010 which was directly attached to the public network operator GDC switch by a 155 Mb/s link.

Paris-Jussieu Ethernet segment was connected to Cisco 7505 Ethernet ATM; ATM AIP 155 Mb/s connected the router to the LightStream 1010 which was connected to a Fore ASX200-BX switch performing VP switching on VPI 5 used between Paris-Jussieu and Orsay-Idris and serving other ATM VP used for native ATM traffic with other destinations. Fore ASX-200BX was connected to the Public Network Operator GDC node.

**Note:** On both LightStream 1010 traffic was paced (shaped) at 10 Mbps, i.e. at the bandwidth set on the ATM VP. By doing this we don't need to control TDP VC (VPI 5 VCI 32) with ctt parameters.

To check the ATM connectivity between both sites, we displayed ATM PNNI adjacency since VCI 18 was active on VPI 5 (but note that PNNI was not available on this VP).

### 4. Results

Connectivity was observed by pinging the interfaces of the routers and switches.

Although we didn't have tools to perform performance measurement, we were able to generate traffic up to 7.7 Mb/s using unsupported TTCP commands on Cisco 7500 (window size = 65535).

We observed a limitation already noticed in phase 1: the impossibility to use another signalling protocol (UNI, PNNI) but the one used by tag switching. For instance we could not set-up a Classical IP on the WAN VP used by tag switching.

As the software already allows to set-up a defined TVC range, tag and other signalling will be probably merged in the same VP in future releases.

This test showed us that it was now possible to test TAG in a full meshed PVP wide area network involving all TF-TEN participants

of this experiment over TF-TEN/James infrastructure.

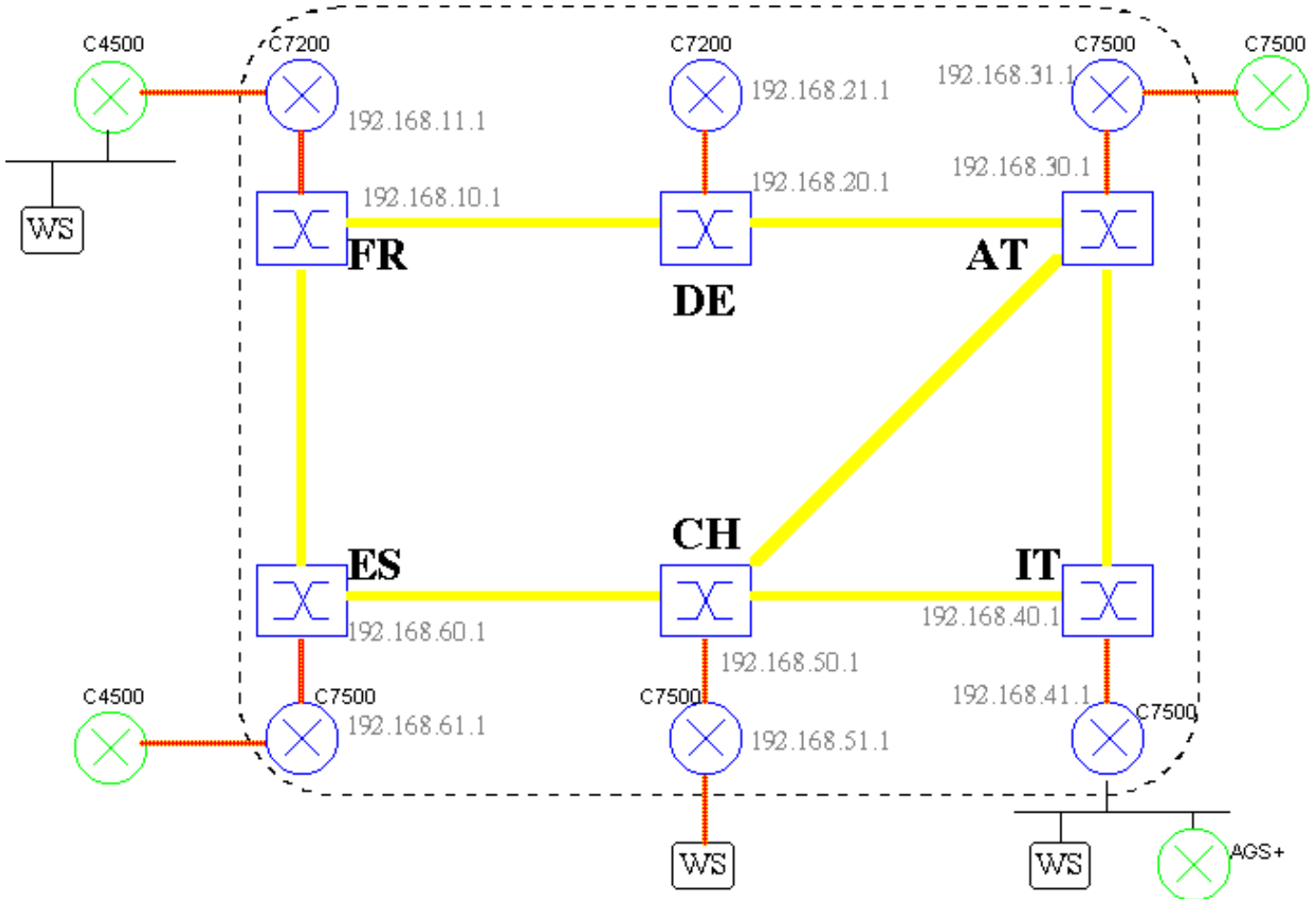
### Phase 3: Experiment on the TF-TEN/JAMES infrastructure

This experiment has been performed by TF-TEN participants on the TF-TEN/JAMES infrastructure from January 1998 the 12<sup>th</sup> to the 23<sup>rd</sup>. European countries involved in this experiment were Austria (ACONET), France (RENATER), Germany (DFN), Italy (GARR), Spain (REDIRIS) and Switzerland (SWITCH).

The goals of this final phase, was:

- to improve the technology in a real large scale network
- to study how this technology can be deployed in a wide area ATM backbone to achieve scalability

#### 1. Testbed



— 2 Mbps ATM VP

— OC-3 ATM link

— Ethernet link

⊠ LS1010

⊡ TAG Switching network (using OSPF)



-----  
**network (using OSPF)**

Figure 3: Phase 3 network infrastructure

The network infrastructure used to perform the tests is the European ATM network JAMES [3]. The overlay network consists of ATM CBR permanent virtual circuits with different capacities depending on the link: 4515 cells/sec on the links France-Spain and France-Germany, and 4750 cells/sec in the rest of the overlay network illustrated in figure 3. The infrastructure consists of a core tag switching cloud (corresponding to the area inside the dashed line) and a set of peripheral non-tag switching local networks in Austria, France, Germany, Italy, Spain and Switzerland. The loop between Austria, Italy and Switzerland was configured to verify the use of the metric for a correct route computation between the three countries.

The tag cloud was made of LightStream1010 switches and CISCO routers of the 7500 and 7200 series, all running tag switching beta software. The ATM switches constitute the core of the tag backbone, because they provide very high performance switching, while routers are deployed on the periphery, because they provide high level routing features necessary to interconnect external networks. In each country we have set-up a tag core switch and a tag edge router. The routers are connected to their adjacent switch with a STM1 link, while switches are connected through the JAMES infrastructure (see above). The tag switching protocol is entirely tunnelled in the JAMES infrastructure, as it was validated in Phase 2. So tag switching was completely transparent for the ATM equipment on the public network operators side.

The tag switching protocol uses an IP routing protocol to exchange all routes through the backbone, in order to set-up the corresponding Tag VCs (TVCs). On each tag equipment an IP loopback address was used by the TAG Distribution Protocol (TDP). Moreover an OSPF routing process was configured in each tag switching apparatus. TDP uses a dedicated control PVC which is automatically configured between adjacent devices (during initialization phase), to exchange all IP routes and establish a full mesh of TAG VCs.

Outside the TAG network, we have connected external networks, i.e. routers (C7500 and C4500) and workstations which were used for the performance tests by generating TCP and UDP memory-to-memory traffic between them. The three workstations (one Sun Ultra in France and Switzerland and a Sparc Station 10 in Italy) were connected to the network through Fast Ethernet, Ethernet and ATM network interface card respectively.

## 2. IP architecture

### for scalability

At the IP level, one solution to interconnect the external networks through the tag backbone could be the redistribution of each external route into the TAG-OSPF protocol. In this way we would have a TVC from each TAG edge router to each external route announced to the backbone, as observed in phase 1. This solution implies a very high number of VCs in the backbone, depending on IP route and tag device numbers, so it does not scale.

Scalability is the capability of the network to cope with an increase in size, complexity, number of users and services offered by keeping at the same time functionality and flexibility.

To achieve scalability, we have set-up a BGP based architecture showed in figure 4.

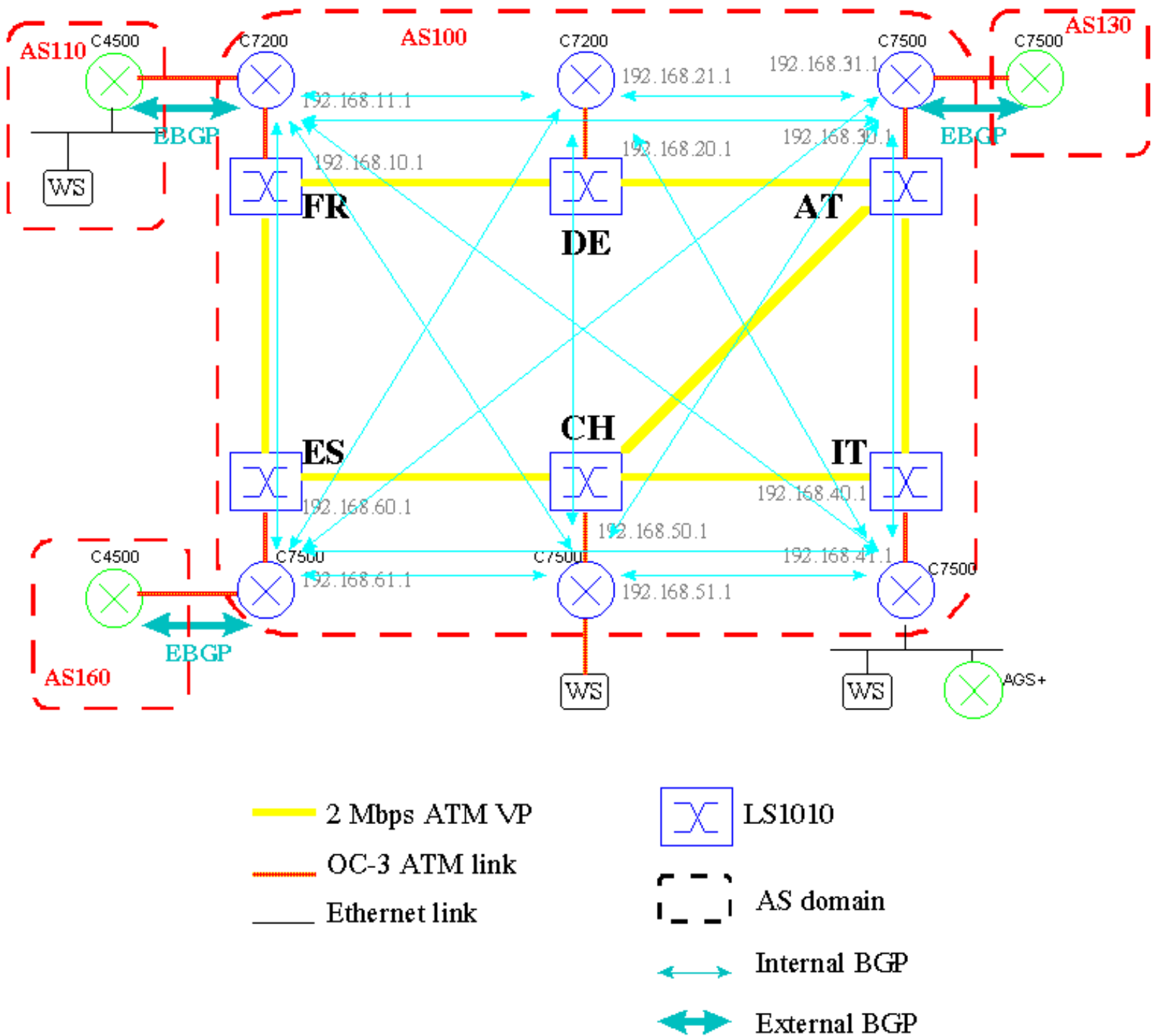


Figure 4: IP infrastructure for scalability

Scalability was achieved through a hierarchical IP routing configuration.

Switches in the core and routers at the edge of the tag cloud run the interior routing protocol OSPF, which is used by the tag switching protocol.

All tag routers (192.168.11.1, 192.168.21.1, 192.168.31.1, 192.168.41.1, 192.168.51.1 and 192.168.61.1) constitute a unique AS backbone and run exterior BGP sessions with exterior routers, whose networks are associated to their own AS number.

In addition, a complete mesh of internal BGP sessions is set-up between tag edge routers. These IBGP sessions rely on tag switching virtual circuits (TVCs) established by tag. These IBGP sessions permit to exchange external routes between tag edge routers.

In this architecture, all the IP datagrams to destinations reachable through a given tag edge router, are forwarded through a unique

TVC. For instance all the traffic between France (AS110) and Spain (AS160) is forwarded through a single TVC set-up between the French tag edge router (192.168.11.1) and the Spanish one (192.168.61.1). The total number of TVCs doesn't depend on the routing table size.

So this architecture permits to have a tag backbone completely independent of the connected networks:

1. The number of TVCs does not depend on the number of external routes: it depends on the number of tag edge routers, tag core switches, physical links (or PVP) and interconnection IP subnets. Estimating the exact number of TVCs in the tag backbone is not easy, but we can say that such a number is proportional to  $n^2$ , where  $n$  is the number of physical nodes in the tag backbone.
2. the tag backbone will not be affected by external routing instabilities. With this configuration, if there is routes flapping on external networks, the tag backbone will not reconfigure the TVCs.

So, through a proper IP routing configuration a scalable tag switching network is achieved.

### 3. Performance

A set of performance tests was done to verify the correct functionality of routing and switching of equipment running tag switching and to compare results achieved in the same network set-up with and without tag switching. Without tag switching tests were performed by configuring static IP routes between the routers directly connected to the switches, in this case static ATM circuits were just used as point-to-point permanent links with RFC1483 IP on ATM encapsulation.

Single and multiple TCP or UDP connections with and without tag switching were set-up between France, Italy and Switzerland. Results obtained with and without tag switching have been compared. *Netperf 2.1* [7] is the application used for measurement of throughput for TCP connections, while *Mgen 3.1* [8] was deployed to generate UDP streams at a given user-supplied rate. The parameters monitored during the test are:

- *round trip time*,
- *route recovery time*,
- *throughput* of TCP connections,
- *packet loss* for UDP streams,
- *average CPU utilization* of routers.

Testing was performed using three end-systems located in Italy (192.168.43.3), Switzerland (192.168.52.1) and France (192.168.17.2) (respectively one Sparc Station 10 and two Sun Ultra, all mounting Solaris 2.5.1). The ATM link capacity of 4750 cells/sec (on the path Italy-Switzerland) gives 1.78 Mbps of application data throughput (considering an IP MTU of 1500 bytes). On the other hand the minimum capacity of 4515 cells/sec (on the path Italy-France) corresponds to 1.68 Mbps of application data. Measuring the real benefit of tag switching in terms of speed in packet forwarding is not easy. First of all, tests should have been done on a loaded backbone, while in our case the PVC bandwidth was not enough to load the network. Then, the IP topology of the testbed is not complex as a real production network, so the IP control protocol overhead is not a critical factor even with a traditional IP architecture. Moreover, in the non-tag switching configuration we just used static IP routes. For this reason, in our testbed the primary goals were the design of the tag switching and IP infrastructure, and the analysis of tag switching functionalities and applicability.

#### 3.1 Round trip time

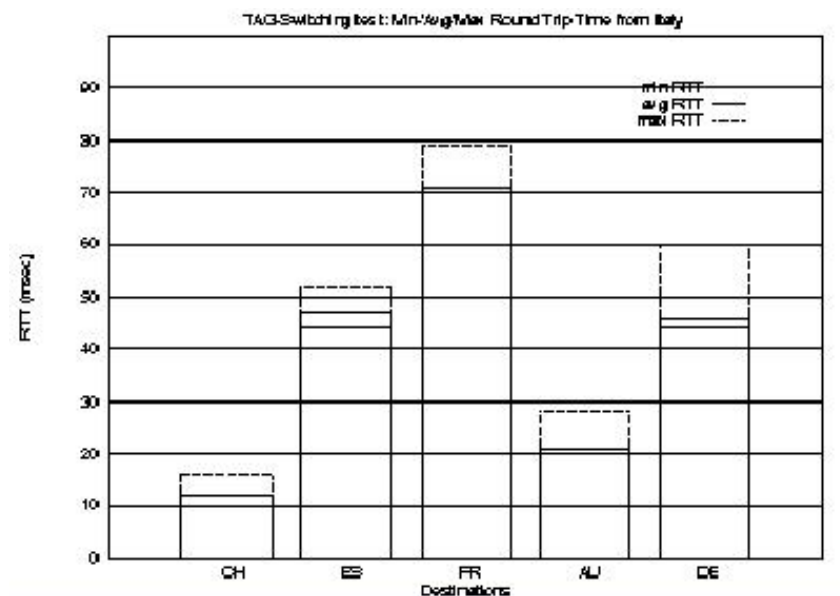


Figure 5: Round trip times between the edge routers of the tag cloud

Picture 5 illustrates the round trip times collected between one of the edge routers (192.168.41.1) to the others. Differences in RTT are mainly due to different propagation delays between the couples of routers. We had the same values both with and without tag switching. This shows that the packet forwarding speed achieved with tag is as good as the one obtained when edge routers are physically adjacent, since the permanent ATM CBR connection is equivalent to a point-to-point physical link.

### 3.2 Route recovery time

One of the goals of the tests was the analysis of the efficacy of the mechanism for route computation of tag switching in an unstable environment. We measured the *route recovery time*, i.e. the time necessary for the router to recompute the path to a given destination when a link failure occurs. During the tests, the link failure event was artificially generated from the switches by repeatedly shutting down the ATM sub-interface corresponding to a given destination. The resulting recovery time is not always the same. The link failure event was generated many times and in different part of the backbone.

The test was very fruitful as we observed a high instability on the tag routers running BGP. The BGP process made the TAG process loop indefinitely during the new tag allocation phase. In this way TDP indefinitely propagates new tags through the backbone, which becomes unstable and unavailable until IP routes are cleared on the concerned tag edge routers. This problem, identified as a bug by CISCO engineers, was reproduced in their laboratory after our experiment and fixed. So we were not able to evaluate the whole recovery time including tag switching recovery time (OSPF) and external network peering recovery time (BGP). The first one (tag switching recovery time) was observed varying between approximately 12 sec and 38 sec.

In spite of this problem, we were able to observe the recovery time for external networks point of view (connected by BGP routing protocol) by reactivating a previously down link. The time needed by the backbone and external routers to recalculate the best routes and establish the right TVCs was measured between 10 seconds and 40 seconds approximately. The recovery time when a link failure occurs is probably longer because it is longer for a routing protocol to detect a loss route than to detect a new one.

### 3.3 TCP Performance with and without tag switching

Tests were done both generating half-duplex connections (i.e. connections with a single source and a single destination) and also with full-duplex streams (i.e. each end-system acts as sender and receiver at the same time). The following paragraphs illustrate the results obtained in the two separate environments.

#### TCP half-duplex connections

Performances have been gathered for concurrent half-duplex TCP flows. Pictures 6 and 7 show the throughput obtained by a single

TCP stream on the path Italy-> Switzerland and France-> Italy respectively. Several values have been collected for different socket buffer sizes (buffer sizes were set consistently on both the sending and receiving side). Results show that in both cases the maximum available capacity on the ATM link is usable. The socket buffer sizes are not relevant in the [first case](#) since the minimum RTT is 12 msec, while [in the second](#) throughput decreases for small socket sizes, because of the *stop and wait* behavior, which is due to the combination of small buffers and of a rather large RTT (70 msec). Because of this bandwidth utilization turns to be rather inefficient.

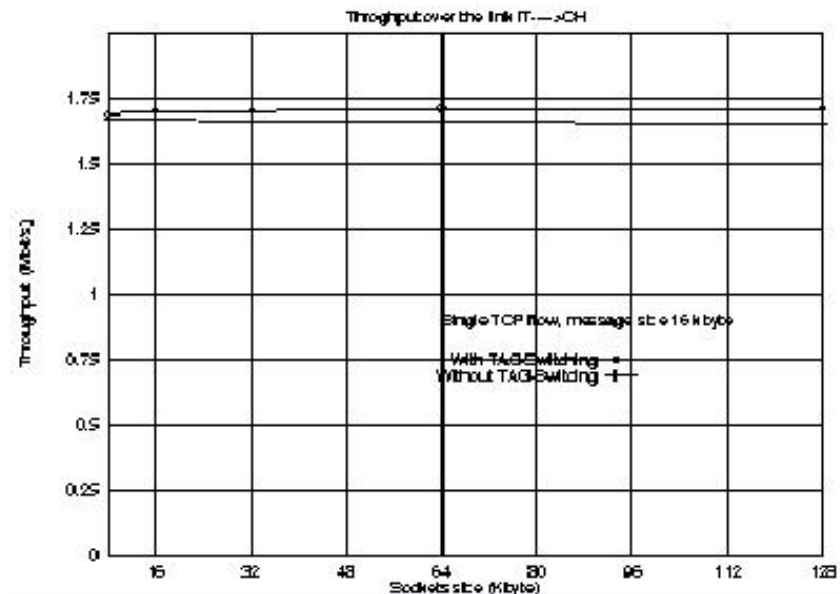


Figure 6: Single half-duplex TCP stream from Italy to Switzerland

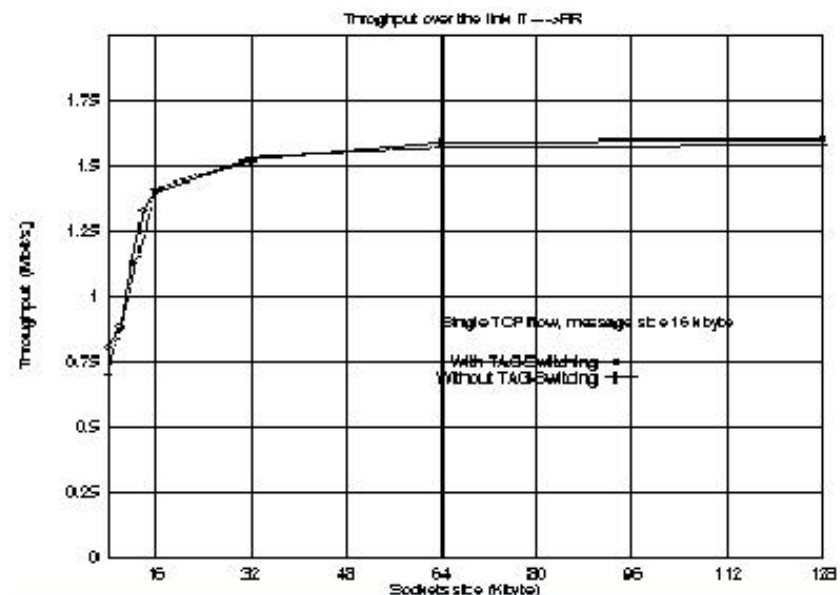


Figure 7: Single half-duplex TCP stream from Italy to France

Results in picture [6](#) and [7](#) refer to tests with message size equal to 16,000 bytes. Tests with different message sizes showed that such a parameter is not relevant in our testbed.

Bandwidth utilization with tag switching is more efficient than without, when ATM permanent virtual circuits and static IP routes are used. In fact, this is shown by the two pictures, which directly compare the two cases. *With* tag switching streams to Switzerland reach 1.75 Mbps against 1.69 Mbps. We had the same for streams from the workstation in Italy to the one in France, in this case *with* tag switching throughput is 1.63 Mbps against 1.57 Mbps. The throughput achieved by connections to France is less than to Switzerland because of the smaller PVC capacity on the path Italy-France. The throughput gain is due to a different encapsulation scheme used in the tag switching test and classical IP test.

TVCs use AAL5 VC based multiplexing encapsulation, while ATM PVCs deploy *AAL5 LLC-SNAP encapsulation* [9]. With LLC-SNAP 8 bytes (LLC header plus SNAP header) are added to the IP PDU when it's encapsulated into the AAL5 CPCS PDU payload. On the other hand, with *VC based multiplexing* no overhead is added at all with a consequent performance gain which depends on the IP PDU size distribution, i.e. on the number of padding bytes added in the AAL5 CPCS PDU.

Note that it is also possible to use VC based multiplexing with IP on PVC, to achieve better throughput.

This set of tests was repeated with even more concurrent connections. The number of flows did not influence the aggregate throughput achieved on the line either with or without tag switching.

### TCP full-duplex connections

Aggregated throughput over the link IT<math>\rightarrow</math>CH with TCP traffic with and without TAG

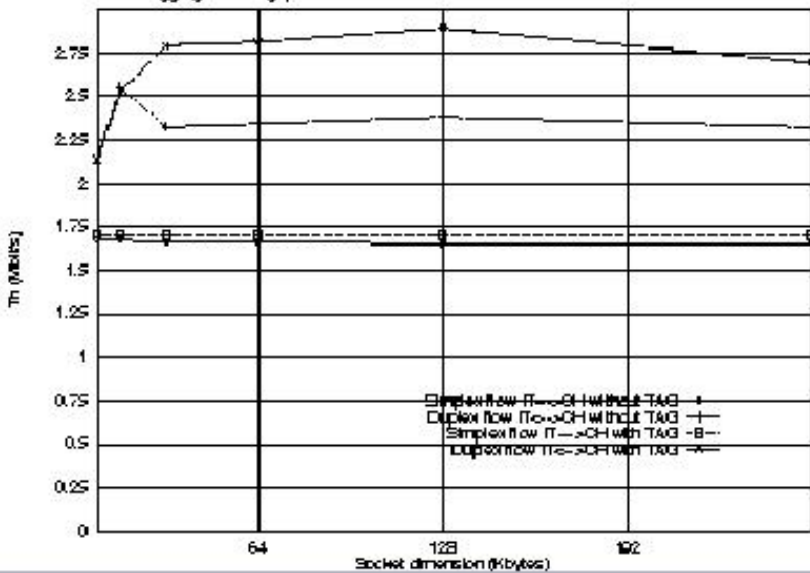


Figure 8: full-duplex connection between Italy and Switzerland

Full-duplex traffic consists of two concurrent TCP connections in opposite directions. The performance test was repeated with and without tag switching and the results are compared in figure 8, which also plots the throughput obtained in both directions for a half-duplex streams to provide a term of comparison.

With full-duplex streams the performances achieved in each direction is less than with half-duplex streams. The performance loss is not constant, it depends on the socket size. The maximum is achieved with socket buffer dimensions around 128 kbytes: aggregate throughput loses approximately 250 kbps. Nevertheless, throughput loss is not

due to tag switching, since full-duplex connections achieve less throughput both with and without tag. Then, with full-duplex streams performance can decrease with small socket buffer sizes.

During the tests we had problems. Throughput achieved in the two directions was not the same, i.e. the direction Italy→Switzerland is less penalised than in the reverse one. Moreover, with full-duplex streams we had high peaks of throughput utilization (an average of 15% in 5 secs) in router 192.168.41.1 (C7200). This problem is probably due to the beta software versions running on the routers, which are still under development, and is argument of future work.

### 3.4 UDP performance tests with and without tag switching

Through Mgen, one or more half-duplex UDP streams can be activated by specifying a given application datagram rate. We used Mgen to produce UDP traffic at increasing data rates between end-systems in France and Italy.

Performances are rather good in the direction Italy → France, since as expected, packet loss starts when the datagram rate overcomes the link capacity. On the other hand, results in the opposite direction France→Italy are not as satisfactory as these, since datagram loss appears with data rates equal or bigger than 0.8 Mbps and the packet loss rate is directly proportional to the application data rate. We think that this is not due to the protocol itself, but to the software preliminary versions running on the routers, especially on C7200. This hypothesis is confirmed by the peaks up to 19% in CPU utilization registered in router 192.168.11.1, while CPU utilization on C7500 was always less than 3%. When repeating the same test between Italy and Switzerland, we had no packet loss in *both* directions for flow rates not exceeding the ATM connection capacity (in this case we just had routers C7500 on the path between the end-nodes). This confirm the correctness of tag switching functionalities in presence of UDP traffic.

## 4. Traffic engineering

*Traffic engineering*, one of the main tag switching features, allows one or more streams, defined through filters, to be forwarded according to a pre-defined path. It gives the opportunity to tailor and balance traffic in the network so that standard routing information can be overridden and well defined streams can be routed differently. The preferential path can be defined as a tag-



switching tunnel. The tunnel is unidirectional and has to be configured in the ingress router only. The rest of the tunnel is automatically and dynamically configured by a signalling derived from RSVP.

The following is an example of tunnel from router 192.168.41.1 to 192.168.31.1. Picture 9 shows the tag switching tunnel (red line) used to route traffic to the Austrian network 192.168.33.0 through Switzerland instead of the direct link Italy-Austria used by the rest of the traffic to the Austrian networks.

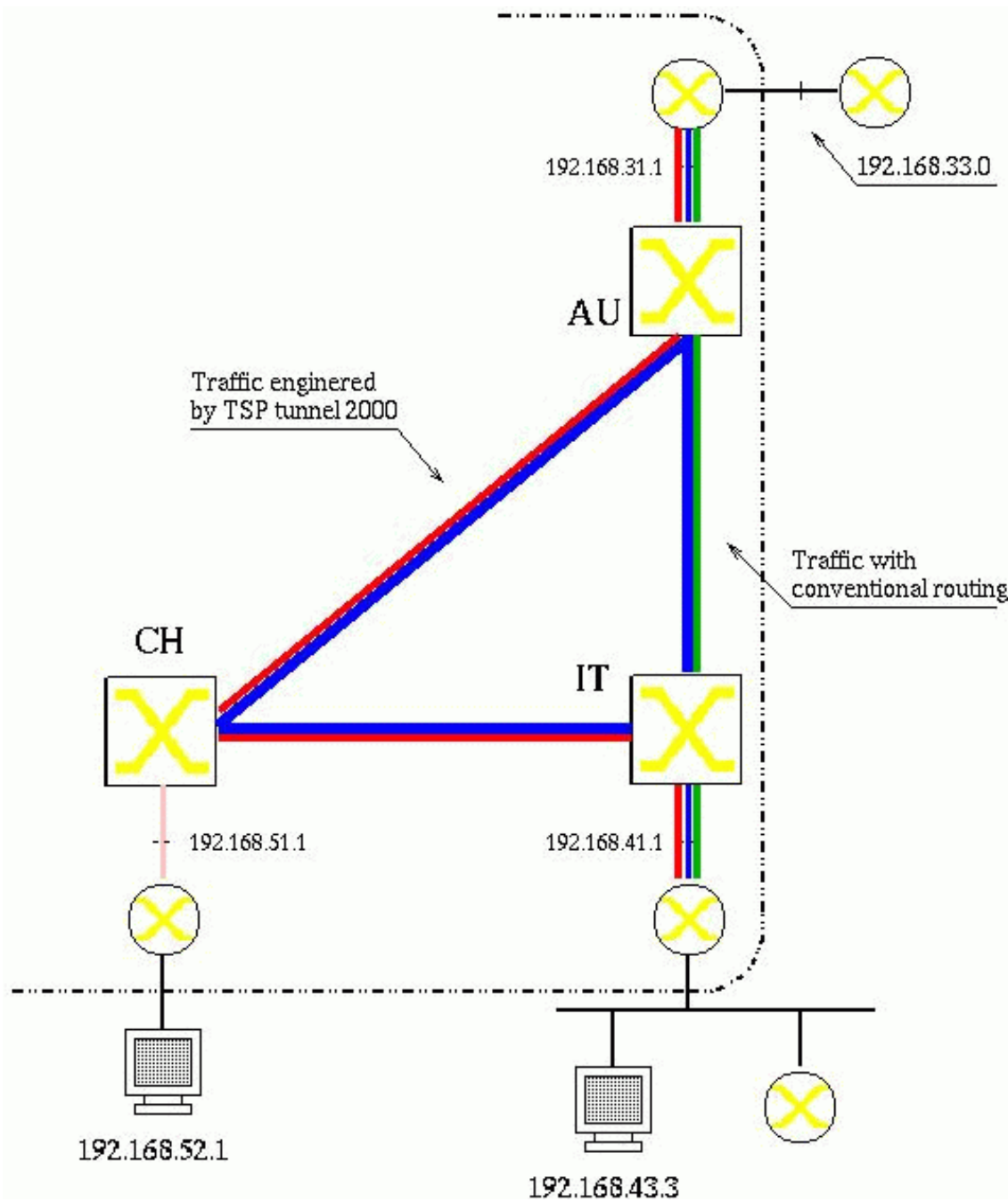


Figure 9: tag switching tunnel (red line) used for traffic engineering

```
interface Tunnel2000
ip unnumbered Loopback0
transmit-buffers
backing-store
tunnel mode tag-
switching
tunnel tsp-hop 1
192.168.40.1
tunnel tsp-hop 2
192.168.50.1
tunnel tsp-hop 3
192.168.30.1
tunnel tsp-hop 4
192.168.31.1 lasthop
```

Traffic engineering works correctly. Preferential traffic to the selected network is routed by overriding the standard route entry:

```
show ip traffic-
engineering
Filter
1: egress
192.168.33.0/24
```

```
Tunnel2000 route
installed
Installed traffic
engineering routes:
Codes: T - traffic
engineered route
T 192.168.33.0/24
(override of routing
table entry) is directly
```

connected, 00:59:30, Tunnel2000

## Conclusions and Future Work

Even if software implementations need to be improved, the tag switching infrastructure in the wide are scale showed good functionality in terms of routing stability, interoperability with ATM, good tunnelling on CBR PVCs, maximum bandwidth utilization with both TCP and UDP and traffic engineering.

The test programme showed that tag switching is a promising and applicable technique.

A more detailed study of performance on a loaded network and the comparison to the results in a traditional equivalent IP infrastructure are two necessary steps for better evaluation.

Also advanced features, as VC merging, VPN, QoS with tag and PIM with multicast tag, not available during the test schedule, but presenting high added value, have to be tested in further experiments. These features and the study of performance are argument of future research.

These tests hopefully encourage people in the MPLS working group to achieve a standardised protocol in a satisfactory time scale. The success of this experiment is also particularly encouraging to face the problem of more and more users and the need of better and more complex services, because through MPLS more scalable CoS-based networks can be designed.

## Bibliography

1. TF-TEN *Label-Based Switching Experiment*, <http://www.renater.fr/jmu/jameslbs.html>
2. *Task Force TEN Homepage*, <http://www.dante.net/tf-ten/>
3. JAMES, *Joint ATM Experiment on European Services*, <http://www.labs.bt.com/profsoc/james/>.
4. R. Callon, P. Doolan, N. Feldman, A. Fredette, G. Swallow, A. Viswanathan, *A Framework for Multiprotocol Label Switching*, Internet Draft, November 1997, [draft-ietf-mpls-framework-02.txt](#)
5. Eric C. Rosen, Arun Viswanathan, Ross Callon, *Multiprotocol Label Switching Architecture*, Internet Draft, Sept 1998, [draft-ietf-mpls-arch-01.txt](#).
6. Cisco Systems, *Scaling the Internet With tag switching*, -white paper- [http://www.cisco.com/warp/public/732/tag/pjtag\\_wp.htm](http://www.cisco.com/warp/public/732/tag/pjtag_wp.htm).
7. *Netperf*, <http://www.cup.hp.com/netperf/DownloadNetperf.html>
8. *Mgen*, <http://tonnant.itd.nrl.navy.mil/ipresearch/mgen.html>
9. Juha Heinanen, *Multiprotocol Encapsulation over ATM Adaptation Layer 5*, RFC 1483, July 1993.



## 4.3 ATM Resource Reservation

### Experiment Leader

Günther Schmittner, Johannes Kepler University, Linz

### Summary

PNNI is an ATM Forum specification for connecting either ATM nodes (switches) or ATM networks. *PNNI* stands for *Private Network-to-Network Interface* and has been approved in its current version PNNI 1.0 in March 1996. It consists of two categories of protocols, one for distributing topology and routing information between physical switches or groups of switches, based on well-known link-state routing techniques, the other for signalling point-to-point and point-to-multipoint connections (SVCs) across an ATM network, based on ATM Forum UNI signalling standards.

PNNI manages and allocates network resources for SVCs in an ATM network. It has to keep track of the current status of all switches and links in order to manage resources in an accurate and efficient way. PNNI has to monitor all allocated resources in the network to be able to decide, whether the requested QoS parameters for a new connection can be satisfied or not.

A PNNI testbed has been set up on the TF-TEN overlay network (see *ATM Routing* experiment). In the final configuration it consisted of 23 switches from 2 different vendors.

The main goal of the experiment was to study how PNNI does resource management, in particular over VP connections (tunnels) provided by JAMES. It turned out, that due to severe hardware and software limits the planned tests could not be done. The main reasons are missing implementations on end systems to specify QoS parameters with certain traffic characteristics and constraints in managing ATM connections inside shaped VP tunnels on ATM switches. Most end systems still request UBR (unspecified bit rate) type service for a connection, which by definition does not guarantee any resources. Therefore it is quite difficult to test resource management in PNNI.

This experiment could be performed in a limited way in the local area, using physical paths between PNNI switches and special software in ATM end systems to create appropriate signalling requests.

### Participants

JKU (Austria), SWITCH (Switzerland), RUS (Germany), RedIRIS (Spain), Renater (France), INFN (Italy), University of Twente (Netherlands), UNINETT (Norway), RCCN (Portugal), UKERNA (UK)

### Goals

- Study operation of resource management in ATM switches running PNNI, especially for the WAN
- Prove interoperability between different PNNI implementations
- Verify correct resource management in PNNI-based networks
- Prove the applicability of PNNI for resource management in a European ATM infrastructure

## Network infrastructure and description

When an end station requests a connection with specific QoS parameters, PNNI is able to find a possible path (if any) satisfying the request and will allocate the necessary resources in the network. It can be mathematically proven that it is not possible to find the optimal path, except by an exhaustive enumeration algorithm, which is not feasible. Some implementations allow a specification of criteria as a base for the selection of possible paths.

Path selection and allocation is done using two different forms of Connection Admission Control (CAC). These are called Generic Connection Admission Control (GCAC) and Actual Connection Admission Control (ACAC). Generic CAC is needed in the path selection process to determine if a link or node is likely to have enough resources available to support the proposed connection. In essence, GCAC predicts the outcome of the actual CAC performed at a switching system.

The experiment planned to study resource management of PNNI in a European test network. As this network was based on tunnels over the JAMES infrastructure (using VP bearer service) and mostly CBR service, it was of particular interest, how different service classes are managed by PNNI.

## Network configurations

Most of the participants were using a Cisco LS1010 switch to connect to the PNNI cloud. At a later stage also FORE switches were integrated (see *ATM Routing* experiment).

Almost all of the public network operators do traffic policing at the entry point to their ATM network (in our case JAMES). Thus we had to make sure, that we did not violate the traffic contract for the 2 Mbps CBR connections, in particular not to send too high bursts of cells into the public network. Therefore we had to implement traffic shaping for the VP tunnels. Unfortunately this is not supported on the Cisco LS1010 platform with the standard feature card (FC-PCQ) of the main processor board (ASP). Traffic shaping is only possible on physical interfaces of the switch using the `atm pacing` interface command.

The Cisco implementation however, supports VP tunnels as logical interfaces on a physical interface, therefore for all partners running more than one tunnel to the PNO we had two choices:

- Use the `atm pacing 2000 force` command on the physical interface. This actually shapes **the sum** of all VPs to 2 Mbps.

- Allocate one physical LS1010 interface per VP and mix them finally together into one outgoing physical link towards the PNO (there was only one JAMES access point of course). In a two switch or loop configuration this actually requires  $2*n+1$  physical ports on the switch, where n is the number of tunnels.

A typical configuration on the LS1010 for the first case is shown here:

```
interface ATM0/0/3
  description Interface towards PNO
  no ip address
  no atm auto-configuration
  no atm address-registration
  no atm ilmi-enable
  atm pacing 2000 force
  atm iisp side user
  atm pvp 4
  atm pvp 5
  atm pvp 6
!
interface ATM0/0/3.4 point-to-point
  description Tunnel to Switch Zuerich
  no atm auto-configuration
  atm nni
!
interface ATM0/0/3.5 point-to-point
  description Tunnel to RUS Stuttgart
  no atm auto-configuration
  atm nni
!
interface ATM0/0/3.6 point-to-point
  description Tunnel to INFN Milano
  no atm auto-configuration
  atm nni
```

To overcome the limitation of traffic shaping for logical interfaces Cisco now offers a new feature card FC-PFQ (per flow queueing), also known as Feature Card III. Among other things it supports shaped VP tunnels for CBR traffic and substitution of other service categories in shaped VP tunnels. Other features include per-virtual circuit or per-virtual path output queueing and logical multicast support (up to 254 leaves per output port, per point-to-multipoint VC), which were required by other experiments.

The FC-PFQ was only officially announced in Q1 1998 and first customer shipment just before end of our experiment. However, Cisco Europe provided 5 test cards as a loan to the group, which were

shipped in March. This gave us only a few days for installation and testing.

## Results and findings

As mentioned earlier the experiment could not be performed due to two main reasons:

1. Applications and driver software on end systems are currently not able to specify arbitrary service categories, traffic characteristics and QoS requirements. In general, only UBR type connections with or even without PCR are supported. Some end systems can signal CBR connections, VBR is very rare.

Therefore it is pretty hard to create appropriate requests for resources in an ATM network. One possibility would be to use a special application, which is able to create and send arbitrary signalling requests, ideally using the ATM Forum Signalling 4.0 specification (see also *ATM Signalling* experiment). Unfortunately to our knowledge such an application does not exist and we had no manpower to write it.

2. The actual setup of the PNNI network made use of CBR tunnels provided by JAMES. A connection request of any service category (CBR, VBR, ABR, UBR) has therefore to be mapped onto such a tunnel. The whole bunch of VCs of (maybe) different type and characteristics has to be appropriately shaped if it were a physical link. PNNI should manage and allocate resources for such a tunnel as it does for physical paths. However, current implementations of switch software place restrictions on this scenario.

On a Cisco LS1010 equipped with FC-PFQ shaped VP tunnels are supported, by default only VCs of type CBR are allowed inside the tunnel. However, it is possible to substitute any other service category for CBR. As an example UBR connections may be allowed inside a CBR tunnel as follows:

```

atm connection-traffic-table-row index 10 cbr pcr 2000
!
interface ATM0/0/3
  description Interface towards PNO
  no ip address
  no atm auto-configuration
  no atm address-registration
  no atm ilmi-enable
  atm iisp side user
atm pvp 4 shaped rx-cttr 10 tx-cttr 10
  atm pvp 5
  atm pvp 6
!
interface ATM0/0/3.4 point-to-point

```

```
description Tunnel to Switch Zuerich
no atm auto-configuration
atm cac service-category cbr deny
atm cac service-category ubr permit
atm nni
```

The `atm pvp 4 shaped` command applies 2 Mbps shaping to the VP tunnel of service category CBR. With `atm cac service-category` statements UBR type connections are substituted for CBR type. Currently there is only **one** type of service category allowed at a time (per tunnel).

A FC-PFQ requires a certain version of ATM switch processor in the LS1010, namely an ASP-B and at least 64 MB DRAM. As an ASP-B was not available at every site, Cisco was asked to provide also processor modules as loan. Johannes Kepler University installed the FC-PFQ on a LS1010 with 32 MB only. This worked for a couple of days without problem, but suddenly all ports of the switch failed. It turned out that installing the old FC-PCQ resolved the problem. Subsequent trials to install the PFQ again failed.

Germany succeeded installing the new feature card, others did not receive or install the card in time.

At last, no tests have been performed. It should be possible to carry out the experiment in the local area, provided one has appropriate test hardware at hand. Also suitable ATM end systems have to be set up, which support UNI 4.0 (that is, all related specifications like Signalling 4.0, ILMI 4.0, Traffic Management 4.0, etc.).

Advanced PNNI features like route selection criteria and background route computation were not studied.

## 4.4 IP Resource Reservation

### Experiment Leader

Simon Leinen, SWITCH, CH

### Introduction

With a backbone service based on ATM, there are two possibilities to provide end-to-end quality-of-service guarantees. The first is to use native ATM functionalities, i.e. SVCs with CBR/VBR/ABR traffic parameters, the second, applicable in an IP over ATM environment, is to offer reservation capabilities through higher layer protocols such as RSVP.

Previous experiments in the TF-TEN framework taught us that the provision of SVCs with traffic parameters other than UBR was very difficult to achieve in our test infrastructure. In addition, there is a lack of applications (in particular IP applications) with the ability of signaling their QoS needs to the ATM layer. For these reasons, we decided to pursue the second option, using RSVP as the signaling protocol for QoS.

The goals of the experiment were to acquire experience with RSVP configuration and use, to analyze the main aspects of its functionality, to test the interoperability of existing implementations on routers and end-systems, and to evaluate RSVP's feasibility for a backbone service. With this in mind, we performed the following tests:

1. Basic functionality

We verified the correctness of the exchange of PATH and RESV messages - through the set-up of reservations without sending real traffic - and of the admission control function by intentional over-subscription of available bandwidth. Moreover, the reservation mechanism was tested with both unicast and multicast routing.

2. Classes of service and reservation styles

The implementation of some more advanced features was tested: controlled-load and guaranteed service, and reservation styles (wildcard filter, fixed-filter, shared-explicit).

3. Reserved flow isolation

We have measured if and how the number and aggregate data rate of best-effort streams can impact service quality for a reserved flow.

4. Protocol scalability

We have measured the impact of the number of reservations on CPU utilisation in the routers along the path.

5. Reservation-based video conferencing tools

A video conferencing application with RSVP support was tested to qualitatively evaluate the efficiency of the controlled-load service.

### Testbed

The test configuration included two 2Mbps CBR VPCs on the JAMES ATM pilot network, one between Bologna (Italy) and Zurich (Switzerland) and another one between Zurich and Utrecht (the Netherlands). The RSVP test network included Cisco 7500 routers in all three locations. IOS 11.2(11)P was used on the routers. In Bologna and Zurich, Sun workstations running Solaris 2.5.1 were used as end-systems (two in Bologna, one in Zurich). The Suns were connected to the routers over 10Mbps Ethernet. In addition, an Intel PC running FreeBSD 2.2.5 in Bologna was used for a later phase of the experiment. The end-system in Utrecht was an Intel PC running Windows 95, connected to the router over an ATM PVC. Unfortunately we didn't succeed in including the Utrecht system in the experiments because the ATM interface on the PC couldn't support multicast, and that was a requirement for the use of VIC, which was the only application supported by each end-system.

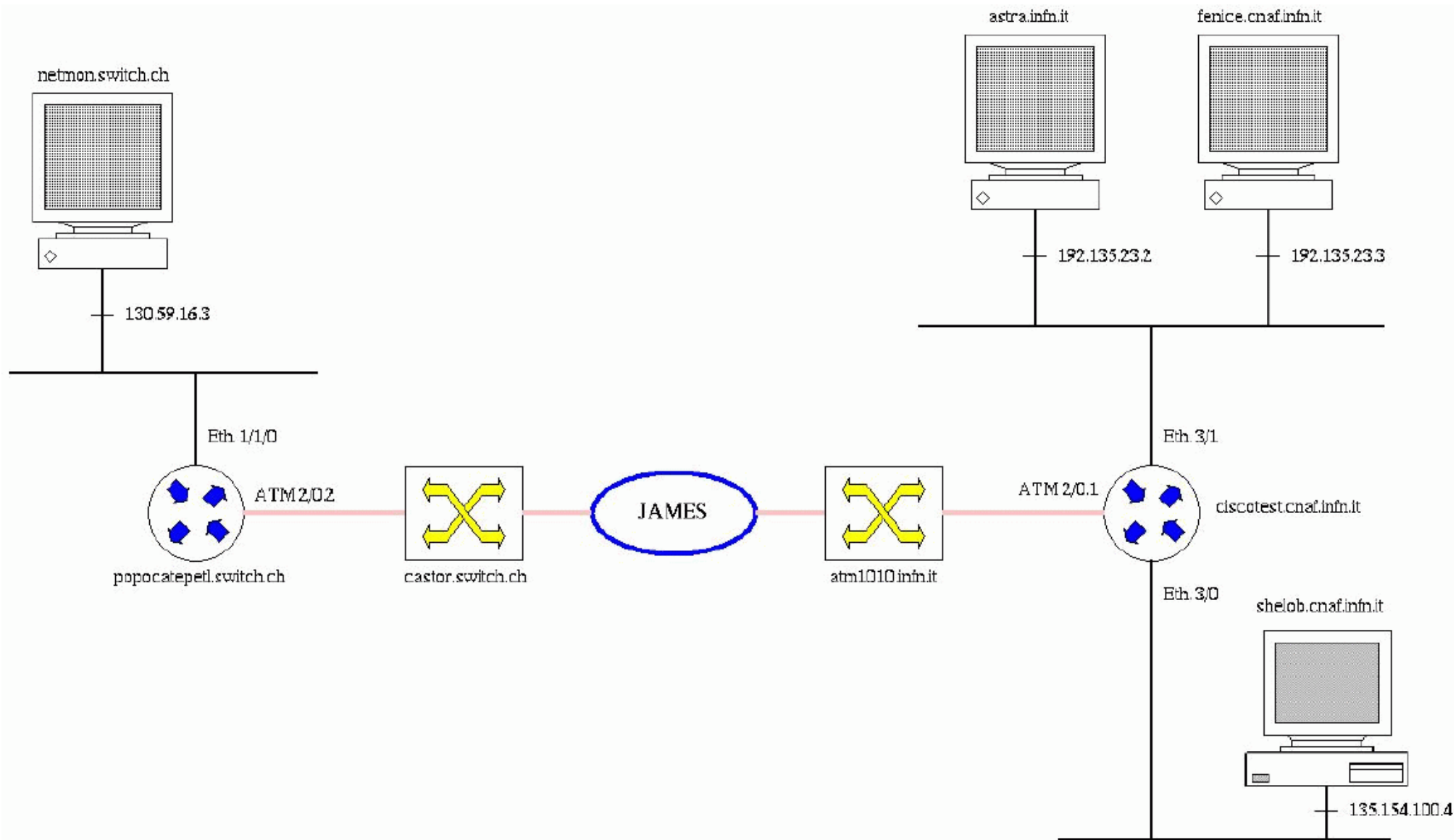


Figure 1

### Basic Functionality

Using the `rtrap` tool, we set up reservations without generating actual traffic, to verify the correct exchange of PATH and RESV messages and interoperability between end-systems and routers. Useful debugging tools were available on both end-systems and routers, and they worked correctly.

On the routers, we could set a limit on the amount of resources that can be reserved on a given link. We used that threshold to test admission control by incrementally activating several reservations until the total amount of reserved bandwidth exceeded the limit. The expected failure messages from admission control were observed. We noted that the thresholds were only taken into account in the "outgoing" directions of the interfaces they were installed on.

### Unicast and multicast routing

Basic functionality was tested with both unicast and multicast routing. For unicast routing, we used static routes to the participating networks. For multicast, the routers were set up for PIM dense-mode routing. The infrastructure is shown in figure 2 below.

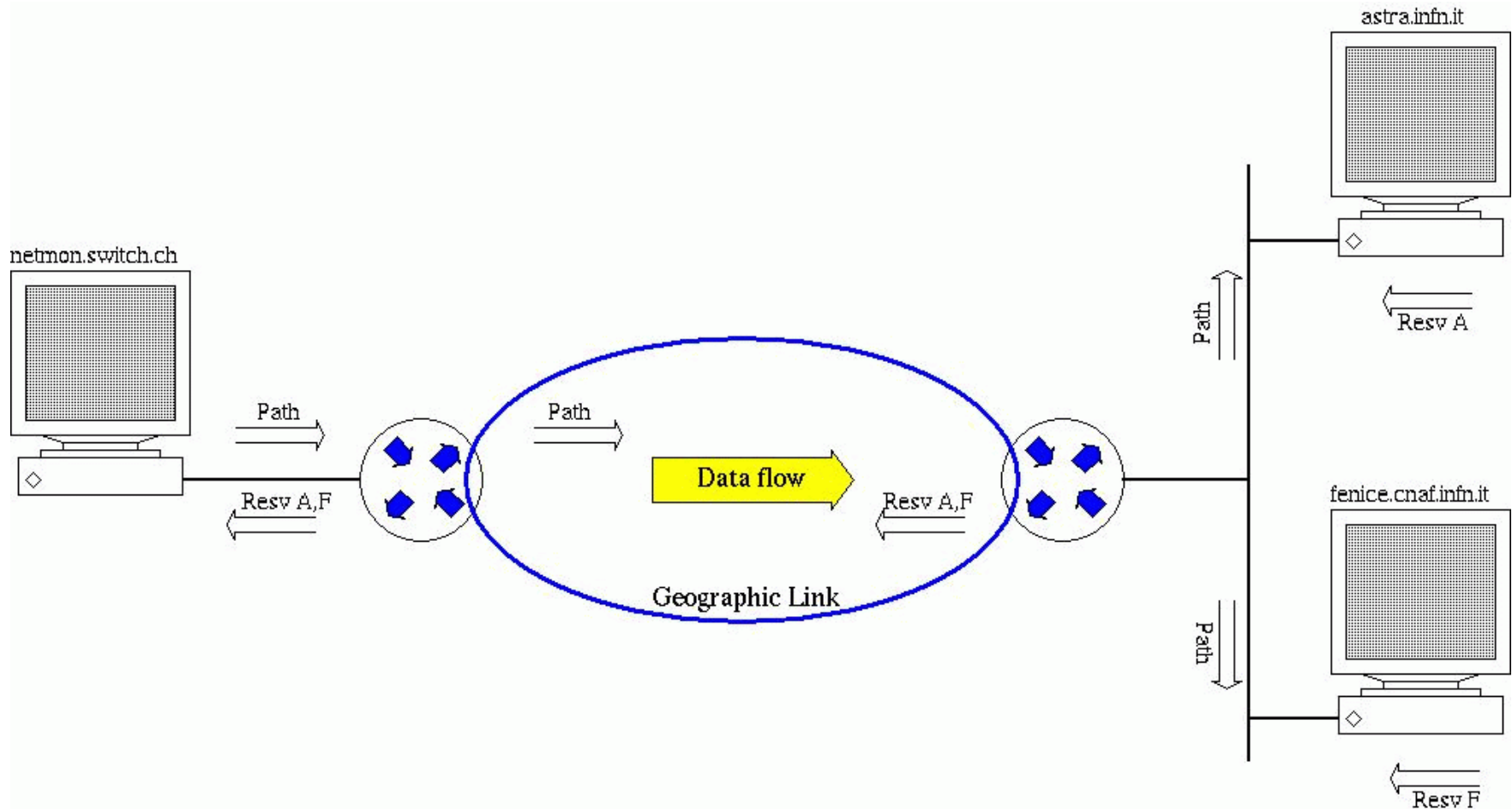


Figure 2

## Classes of service and reservation styles

### Classes of service

We tested reservation setup with both controlled-load and guaranteed service traffic classes, which were both honored correctly on the routers. For the performance tests, only controlled-load reservations were used because no application support for guaranteed service was available on end-systems.

### Reservation Styles

We tested the three reservation styles supported by RSVP: wildcard filter, fixed-filter, and shared-explicit, and all of them worked correctly. The tests included two or three senders and a single remote receiver. The senders send PATH messages with different bandwidths downstream. The use of different reservation styles can be monitored by the router.



To test wildcard filter, we generated a multicast session consisting of three members: one sender and two receivers. When the first receiver R1 joins, it specifies a given amount of bandwidth X (the service under test was controlled-load). If the second receiver R2 joins afterwards and specifies a bandwidth Y bigger than X, then the reservation established on the path is updated, since according to the specifications, the maximum of X and Y is computed and chosen as new reservation. Similarly, if host R2 leaves the multicast group, the current reservation value is refreshed and set back to X.

To test the fixed-filter style, we set up two streams between a single sender and two remote end-systems. With fixed-filter, each receiver specifies the source to which the reservation must be set up. As a consequence, each router on the path indicates the amount of resources allocated by providing the mapping between bandwidth allocated and the couple (sender, receiver). If the sum of resources specified by the receivers exceeds at least one reservation threshold set on the routers on the path, an admission failure is generated. In the CISCO routers a single threshold per router interface can be set. When RESV messages are generated, the threshold which is taken into account for the control is the output interface, where "input" is the interface from which the data traffic is received and "output" is the one to which the data is forwarded.

When a shared-explicit reservation is used, a single RESV message is sent upstream to the group of sources. The RESV message is then copied and sent to each source by the router in which the path between the receiver and the corresponding sources of a multicast group splits into several branches. The following is an extract of the debugging output obtained for a multicast session (224.225.0.1) between one sender (130.59.16.3) and two receivers applying for a shared explicit reservation. "show ip rsvp sender" provides information about PATH messages, "show ip rsvp reservation" shows the reservation requests received from downstream, while "show ip rsvp request" indicates the amount of resources reserved upstream.

```
router> show ip rsvp sender
To          From          Pro   Prev Hop      I/F      BPS
224.225.0.1 130.59.16.3   UDP   195.176.0.25 AT2/0    1500K

router> show ip rsvp reservation
To          From          Pro   Prev Hop      I/F      Fi      Serv   BPS
224.225.0.1 130.59.16.3   UDP   192.135.23.2 Et3/1     SE     LOAD   800K
224.225.0.1 130.59.16.3   UDP   192.135.23.3 Et3/1     SE     LOAD   80K

router> show ip rsvp request
To          From          Pro   Prev Hop      I/F      Fi      Serv   BPS
224.225.0.1 130.59.16.3   UDP   195.176.0.25 AT2/0     SE     LOAD   800K
```

Note that the two reservations have been merged in such a way that a single reservation is sent upstream, with the bandwidth of this request being the maximum of the bandwidths in the two reservations received from downstream.

## Reserved Flow Isolation

### Reserved flow in Local Area Network

In this setup, the two workstations were attached to separate Ethernets which were connected to each other by a router. We generated one reserved flow of 1 Mbps between the two end-systems. Over the same link, a third end-system was used as a source of a single best-effort (UDP) flow of increasing data-rate. Figure 3 plots the packet-loss for these flows.

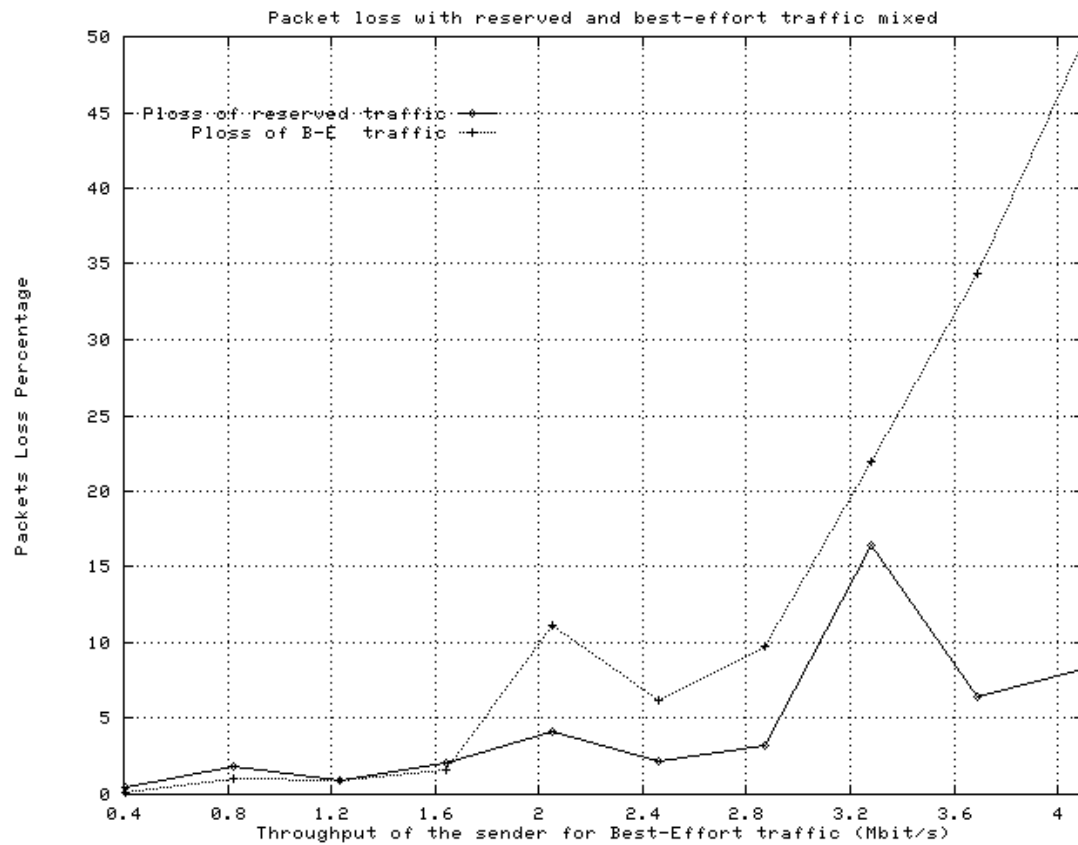


Figure 3

It can be seen that the reserved flow is indeed handled preferentially compared to the best-effort flow, but the impact of the best-effort load on the service quality experienced by the reserved flow was still quite noticeable. The controlled-load traffic packet loss is rather unpredictable: packet loss increases up to 16.4 % when the best-effort data rate is equal to 3.3 Mbps, then it drops. The understanding of this behaviour requires further analysis.

### Reserved flow in Wide Area Network

We repeated the previous test in a wide-area environment. The workstation in Switzerland sent 1 Mbps of reserved data to the NetBSD machine in Italy. This flow competes for use of the resources with an increasing data-rate best-effort flow.

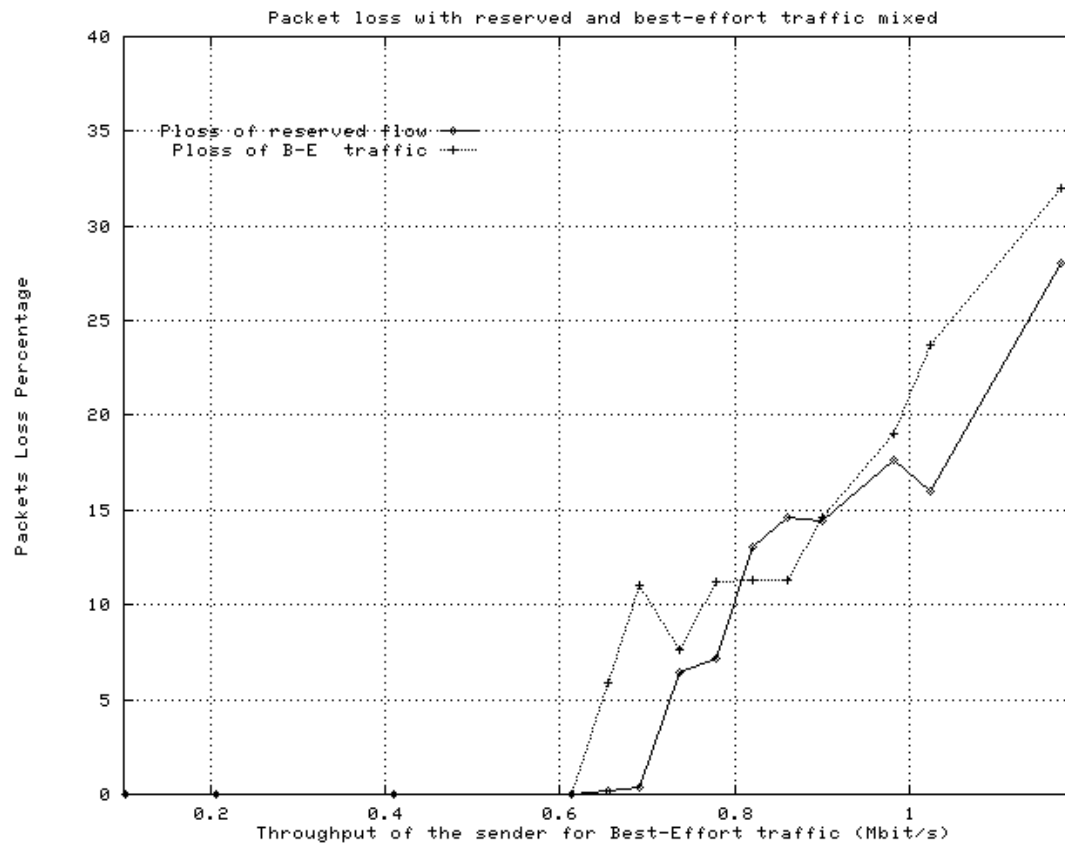
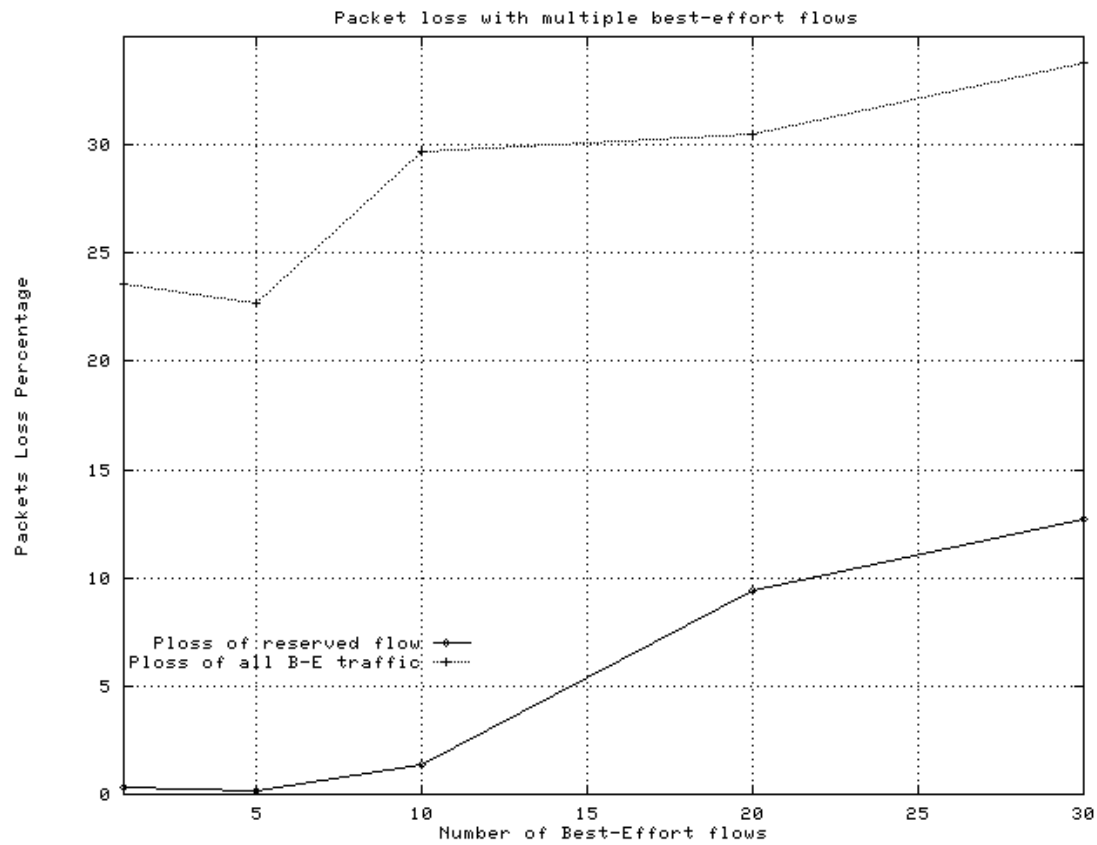


Figure 4

The packet loss of both flows increases as the link starts to become congested (1.6 Mbps of aggregated throughput), but the reserved flow is less penalised. However, the controlled-load service should provide little packet loss also with overloaded network, but the result of this test does not comply with this specification. Furthermore, in some cases the packet loss of the controlled-load stream is bigger than the packet loss for the best-effort stream, this implies that the two types of services are not correctly managed. This could be due to some misconfiguration in the end-systems or to the implementation of CBQ on the Solaris platforms. The reason why this problem arose only in the wide area testbed is argument of future research.

Packet loss of the reserved flow also depends on the number of best-effort flows. We have generated a reserved flow (1 Mbps) and a variable number of best-effort flows in the range [1, 30] for 1 Mbps of aggregated throughput. The packet loss experienced by the reserved flow is smaller than that experienced by the best-effort traffic, but increases with the number of different flows.



### Interaction with best-effort TCP traffic

The increasing data-rate reserved flow competes for the use of the resource with a TCP/IP best-effort traffic generated using *Netperf* [2]. Figure 6 plots the throughput achieved by the both flows as the sending rate on the reserved flow is increased.

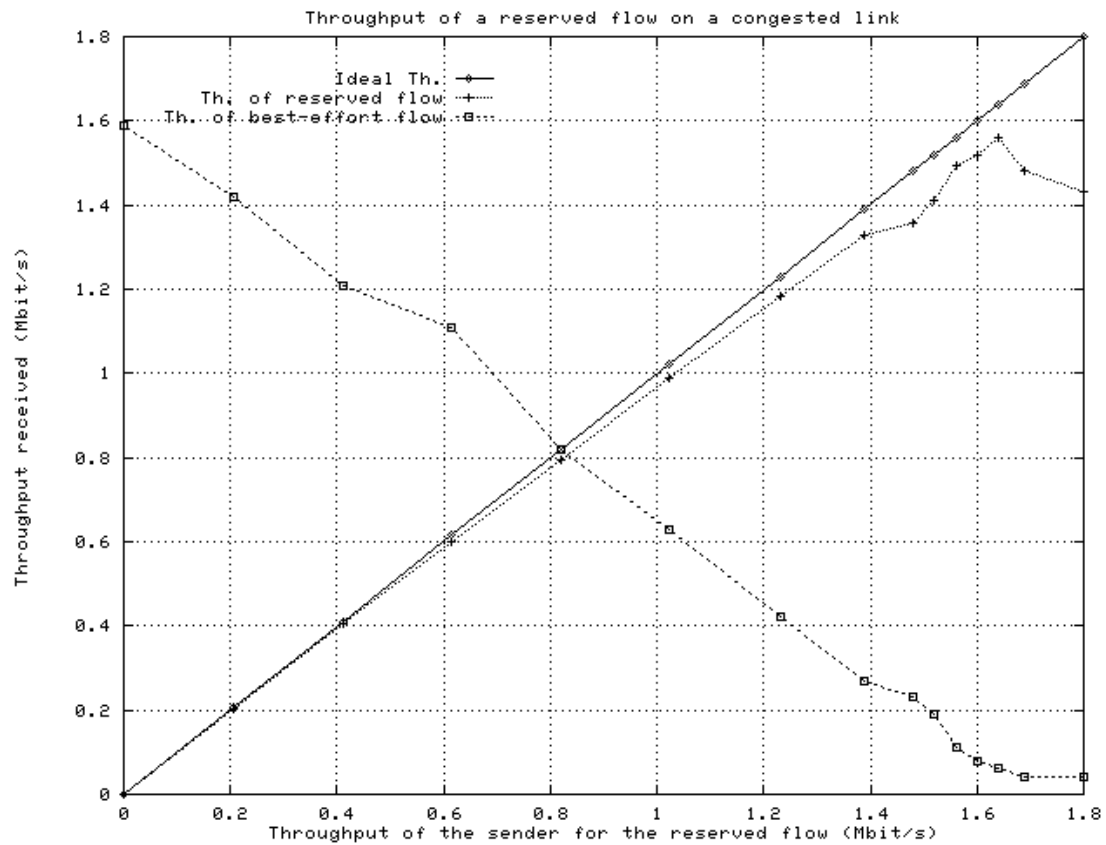


Figure 6

The continuous straight line represents the ideal behaviour. The dotted line plots the throughput of the reserved flow, which is not too different from the ideal behaviour up to a sent data-rate of 1.5/1.6 Mbps, where the router limits the maximum reservable resource, and the traffic in excess is delivered with best-effort service. The dashed line plots the best-effort throughput. The TCP traffic decreases linearly as the UDP reserved flow traffic increases. If the best-effort traffic is TCP, the influence on reserved flows is notably weaker than if it is UDP.

### Protocol scalability

To improve the scalability characteristic of RSVP is necessary to open a great number of session at the same time. The monitoring of the router interfaces state show how and when the reservation are really installed, while the CPU use can show how the process resource are charged.

There is a software limitation of 50 independent reservations per host. As shown in figure 7, we set up one hundred remote and fifty local reservations.

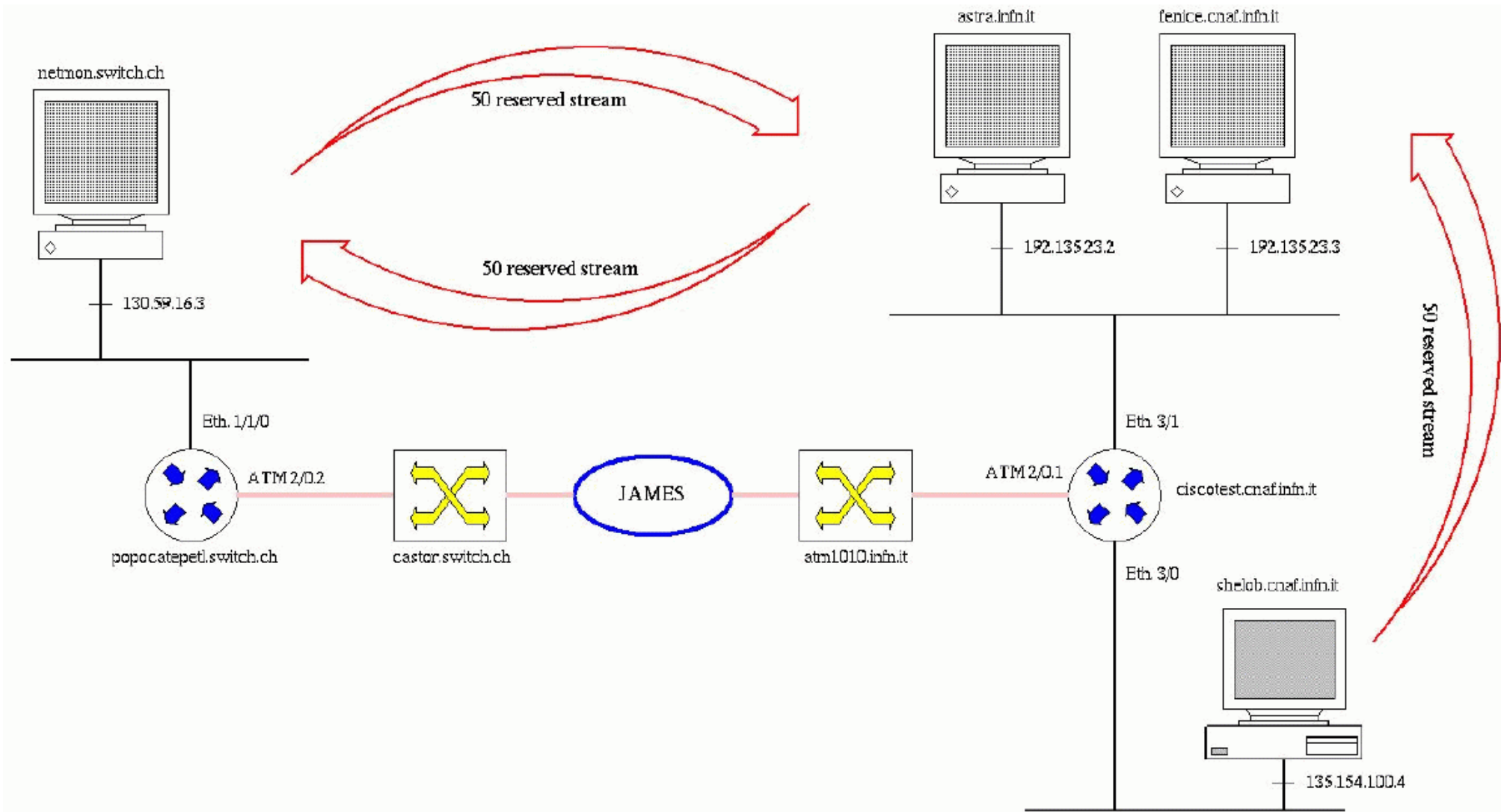


Figure 7

The CPU usage on the router in Italy - which handled most of the reservations - shows peaks up to 4%, with the average value being 2%. Because no user traffic was generated in this experiment, CPU use can be attributed mainly to the signaling traffic.

In conclusion, it seems that the processing load due to reservation handling on routers could be sustainable in a backbone network, provided that the number of RSVP reservations remains moderate. A number of a few hundred active reservations on a network of the size of TEN-34 seems feasible with current equipment.

### Reservation-based video conferencing tools

We tried to understand the efficiency of controlled-load service with a real application, the *vic* video conferencing programme. In order to do that, we ran a remote video conference session between a workstation in Switzerland and another in Italy, and we monitored packet loss using the diagnostics tools built into the *vic* programme. In the presence of best-effort (TCP) background traffic, *vic* packet

loss was measured both with and without a reservation for the video stream.

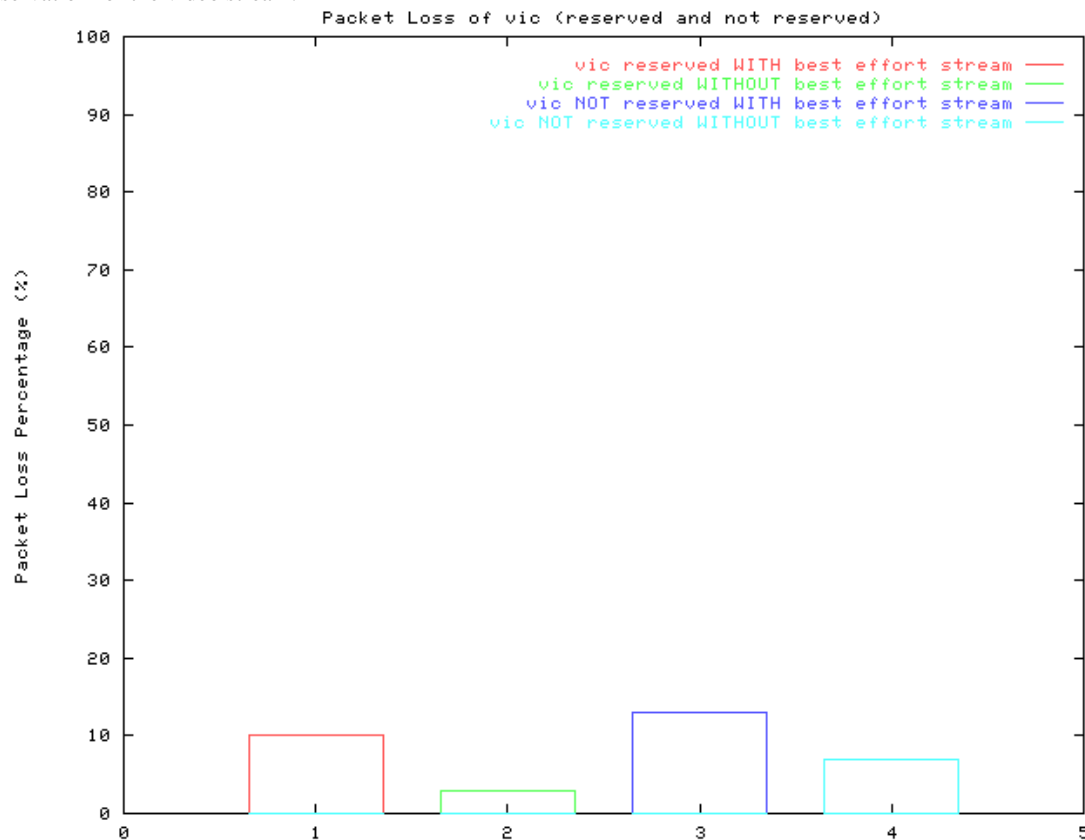


Figure 8

In picture 8 each column corresponds to a different test. Column 1 reports on the packet loss experienced by a controlled-load vic stream with a best-effort run in parallel. As a term of comparison, column 2 shows the performance of the same reserved vic session run without best-effort traffic in the background. Similarly, column 3 and 4 report on the results achieved in similar conditions by a best-effort vic video stream.

Some of the packet loss experienced in this experiment actually seems to be related to processing load by the conferencing application on the end-systems. This made it very difficult to generate reproducible results. With best-effort traffic and a controlled-load vic session, we noticed slightly better performance. A more detailed study of RSVP and vic traffic is subject of future investigation.

## Conclusions

RSVP seems viable and mature as a reservation signaling protocol for IP. The core protocol specification is now stable and has been published as RFCs ("Requests For Comment", in fact standards documents with relatively strong influence on the business community) by the IETF. Interoperability has been demonstrated between several different implementations. The respective leading vendors of routers, PC operating systems, and technical workstations have either announced implementations or are already shipping them commercially.

## Open Issues

However, there are a few important issues which still work against the adoption of RSVP on the Internet or parts of it:

## Mappings to Specific Lower Layers

The realisation of IETF Integrated Services traffic classes is relatively straightforward on transport media with stable and well-defined characteristics, such as PDH or SDH leased lines, permanent CBR/

VBR ATM connections, or point-to-point Ethernet. More complex transport layers such as ATM or Frame Relay networks with SVCs or less well-defined traffic classes, shared LANs, or low-bandwidth modem/ISDN connections require specific mechanisms to ensure maintenance of service quality for reserved traffic. The definition of such mechanisms is ongoing work in the IETF "ISSLL" (Integrated Services over Specific Link Layers) working group. While the standards for ATM, IEEE 802.x LANs, and low-speed serial links are near completion, it is unclear when they will be implemented by vendors of networking equipment.

## Queuing Mechanisms

Aside from transport-layer support, the implementation of Integrated Services requires more sophisticated *queuing algorithms* in routers than are commonly used today.

Today's routers mostly implement a simple "first-in, first-out" (FIFO) queuing discipline, with arriving packets being dropped when the queue is full. Some refine this by adding a special queue for high-priority ("network control") traffic to allow for router management during periods of extreme congestion. There is now a trend towards better queue management using Random Early Detection (RED), which avoids bursts of dropped packets by randomly dropping packets from queues that are filling, but not completely full yet. However, the basic variant of this still implements a strict FIFO discipline.

For effective implementation of Integrated Services traffic classes, routers more sophisticated queuing disciplines involving something like classes or priorities. Class-Based Queuing, Weighted Fair Queuing (WFQ), and Weighted RED (WRED) are techniques for doing this. Choosing between them is a tradeoff between performance (computational cost, particularly at high forwarding rates), code complexity, and traffic isolation.

This phase of the experiment didn't leave us enough time to test the behaviour of different queuing schemes. The results for the (default) one we tested was not satisfying in the sense that ATM-like quality of service guarantees could not be obtained. However, there is no reason why that should not be possible, even with current hardware, by selecting the right queue management scheme.

## Policy Control

As specified, RSVP only performs admission control for new reservations on the basis of available resources. In most cases, one would like to add additional administrative (policy) constraints to the process of determining whether a reservation should be granted. The same problem exists for ATM networks with signaling, and in fact there is a TF-TEN experiment treating this issue in a way that applies to both RSVP and ATM.

Another part of this domain is the enforcement of the traffic contracts negotiated by RSVP users, i.e. the policing of RSVP flows against their announced bandwidth/burst size characteristics. This should be tested more thoroughly in the future.

## Business Models and Inter-ISP Issues

The Internet consists of thousands of independent Internet Service Providers' (ISP) networks, interconnected in a complicated mesh. The contracts that rule these interconnections are usually very simple. Many interconnections today don't even involve settlements, or are billed independently of the amount of traffic transferred.

It is conceivable that a commercial ISP offers reservation capabilities to their customers, presumably for an extra fee (that may depend on the number and size of reservations). But in order for such reservations to be effective beyond the boundary of this particular ISP, its neighbor ISPs (peers, providers, customers) would have to honor reservations from the first ISP. This would involve complex Service-Level Agreements and probably a kind of inter-provider accounting and charging that is completely inexistent in today's Internet (although familiar in the classical Telco world). It is unclear whether the Internet will evolve into a system that would make this work, and if so, how long it will take.

The newly chartered Differentiated Services (diffserv) working group in the IETF [\[3\]](#) is trying to define QoS features for IP which are based on simple per-packet classification tokens. While those services might not allow for strict QoS guarantees like in ATM or RSVP, the necessary SLAs between providers could be much simpler, and the economic incentives of offering this kind of differentiated services may eventually be interesting enough for this to be adopted by a large part of the ISPs that form the Internet.

## Recommendations

Transport-independence and support on a wide variety of end-systems make RSVP an attractive mechanism for the provision of end-to-end QoS, provided it is used in the right context:

- In the absence of an elaborate policy control infrastructure, its use should be restricted so that accidental or intentional misuse by few users doesn't deny service to many.
- In the absence of elaborate accounting and charging, use of the service must be controlled by weaker means such as user education, trust, and mutual agreement.
- With given hardware, the number of reservations that can be sustained will always be of a smaller order of magnitude than the number of best-effort flows supported by current IP backbones. This



makes the service uninteresting for mass applications such as generalized Internet telephony with a reservation for each call. However, other applications such as teleteaching or remote diagnostics could be supported by Internet backbones at relatively low cost.

Campus networks and national research networks (NRNs) are natural candidates for RSVP deployment. Individual commercial ISPs may find it interesting to offer RSVP service to their customers, but it is hard to imagine how RSVP-like QoS guarantees can be offered over ISP boundaries in a commercial setting.

A backbone network connecting NRNs could provide much added value to the user community by supporting RSVP. To this end, a management concept would need to be developed that allows limited use of this service while preventing waste of resources due to excessive reservations. Such a concept could be conceived as either a centralized model where every use of RSVP would have to be permitted by a central (network) management entity, or a more distributed one where a-priori permission is granted to different participants, who in turn take on responsibility for the reasonable use by their respective community.

## References

1. R. Braden, Ed., L. Zhang, S. Berson, S. Herzog, S. Jamin, [Resource ReSerVation Protocol \(RSVP\) -- Version 1 Functional Specification](#), RFC 2205, September 1997
2. R. Jones, [Netperf homepage](#)
3. IETF Differentiated Services Working Group, [homepage](#)

## 4.5 ATM Point to Multipoint

### Experiment Leader

Robert Stoy, Rechenzentrum Universität Stuttgart (RUS)

### Summary

The experiment tested the suitability of the point to multipoint connection capabilities through WAN-ATM networks available today for group communications in the special case of IP multicast on top of ATM point to multipoint SVCs. On top of the SVC TF-TEN network the Protocol Independent Multicast Sparse Mode [1] (PIM SM) protocol was used between multicast routers.

The basic result is, that in the part of the TF-TEN ATM SVC network used the ATM point to multipoint SVCs worked well in conjunction with PIM SM between the routers on the edges of the network. Also that two classes of SVCs are used, first point to point SVCs between the multicast routers for transmitting the PIM protocol data units, and second a dedicated point to multipoint SVC tree for each group session. During the tests only UBR service was available on the SVCs.

The ability of the network equipment of serving other service classes than UBR in conjunction with point to multipoint SVCs is for further investigation.

### Participants

DFN/RUS (Germany) , AConet/JKU (Austria)

### Dates and phases

#### *Phase 0: local p2mp PVC tests*

Local preliminary tests with static p2mp configurations were done until November '97.

#### *Phase 1: local tests with SVCs*

The tests with p2mp SVCs were divided into two phases. In the local test phase beginning in Jan '98 a suitable router and switch configuration was tested in a two node configuration.

#### *Phase 2: PIM SM SVC tests through TF-TEN network*

In the second phase during March '98 the configuration was extended through the TF-TEN SVC network with a third node in Austria.

### Goals

The main goal was - after an evaluation of ATM Multicast capabilities in ATM networks - the test of the proper function of PIM Sparse Mode protocol on top of the point to multipoint capable SVC ATM network. As an application the mbone tools were used on the endsystems which were connected to the routers on the

edges of the ATM network.

## **Network infrastructure and configuration**

The following figure shows the network infrastructure used for the last test phase with three nodes, two of which are located in Germany and one in Austria. On the network side each node consists of a router which is connected through an ATM switch to the TF-TEN SVC network.

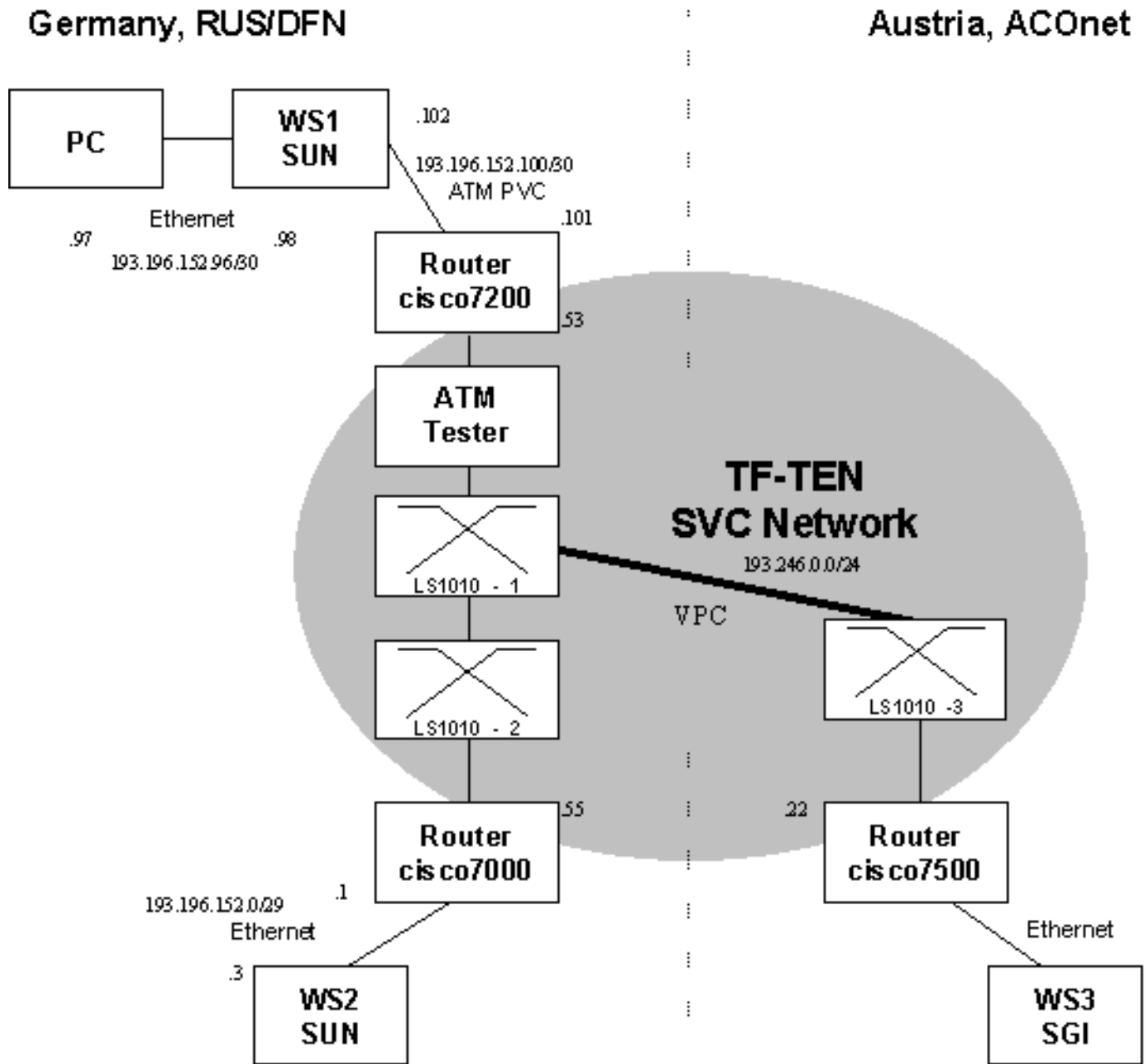
The SVCs between Austria and Germany were tunneled through a VP. The CISCO ATM switches LS1010-1 and LS1010-3 that were connected to the VP DE-AT were equipped with the second generation switching processor (ASP-B with FC-PFQ) during test phase 2.

As edge devices on the TF-TEN SVC network CISCO routers were used. One router in Germany was configured as PIM rendezvous-point router, whereas the other ones had the role of leaf routers.

Two endsystems (WS2 and WS3) running the mbone tools were connected directly to the routers. One workstation (WS1) was connected via a PVC to its router and was therefore configured as mbone router running mouted. Because this machine had no working video/audio equipment a PC with suitable equipment was connected to WS1.

The overall configuration was fully separated from the public MBONE.

An ATM tester was used for monitoring and decoding the signalling messages between the rendezvous point router and the TF-TEN ATM WAN access switch (LS1010-2).



**Figure:** Network configuration in the test session Germany –Austria during test phase 2

The following table shows the SW releases of the network equipment used during the phase 2 test.

c7000	LS1010-1	LS1010-2	LS1010-3	c7200	c7500	WS1
IOS 11.2 (11)	IOS 11.3 PNNI beta release	IOS 11.2 (8)	IOS 11.3 PNNI beta release	IOS 11.3 (2)	IOS 11.3(2)	Solaris 2.5.1 mouted 3.8

## Test description

The tests were started by announcing a multicast session with video and audio on one endsystem at RUS (PC). On the other endsystems it was first checked if the session was advertised via PIM SM and IGMP. Next by starting the video and audio tools and joining the advertised session the transmission functionality on the point to multipoint SVC tree was checked. In order to find results about the proper function of PIM SM and IGMP on the routers the relevant caches were observed and the protocol sequences were monitored. The point to multipoint signalling on ATM level was monitored with the ATM tester shown in the figure above.

## Results and findings

### P2MP capability of ATM switches

Early local testphases showed that the setup of point to multipoint PVCs is not possible between ATM VP endpoints on the equipment used. The upgraded HW (next generation Switch processors on LS1010s) support these functionality as described in the documentation. However this feature has not yet been tested.

### Two types of SVCs

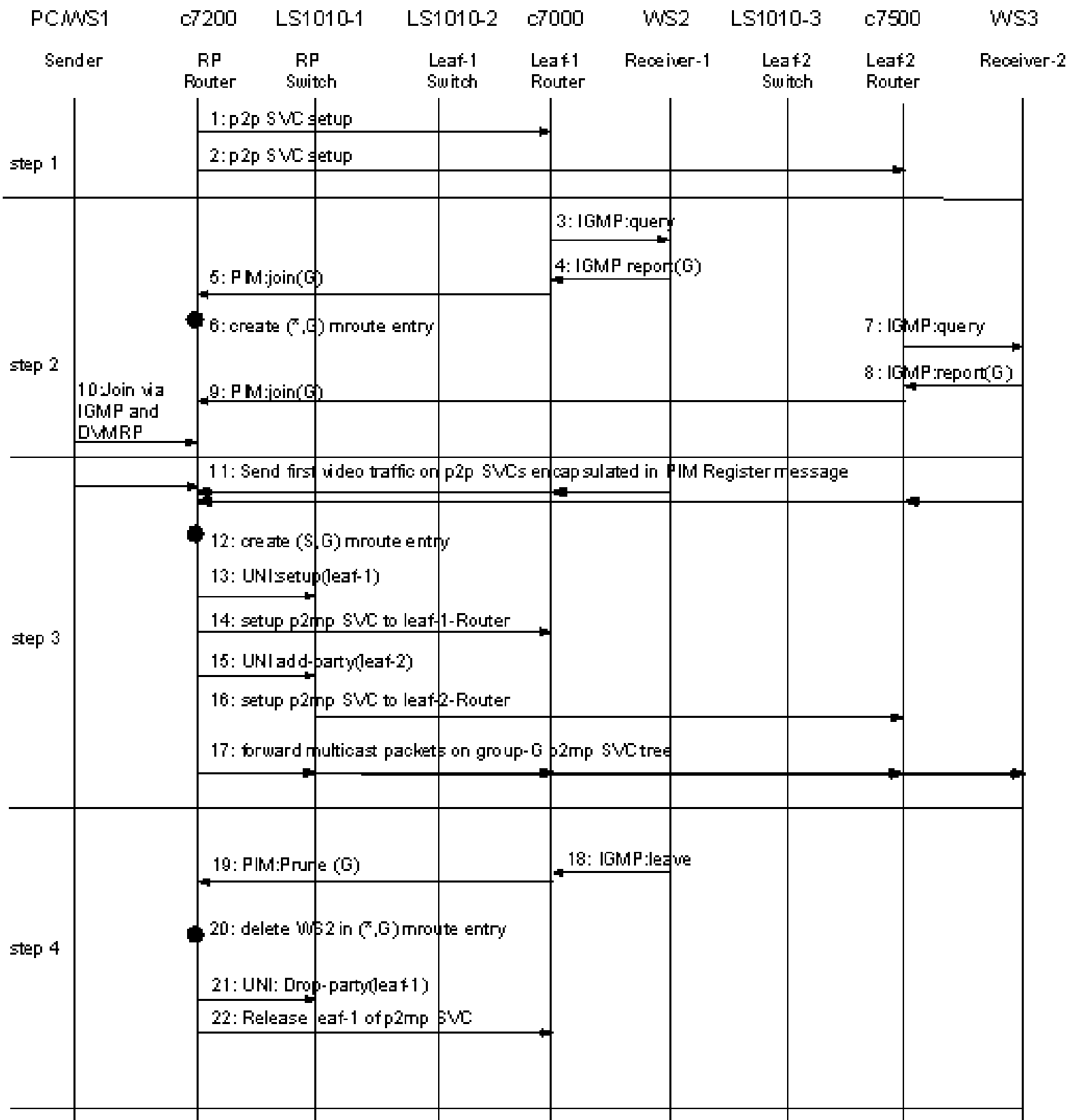
The PIM SM protocol data units are exchanged through a mesh of bidirectional unicast SVCs between an rendezvous point (RP) router and leaf routers. The setup of these SVCs is initialised by PIM itself. For the transmission of application data streams (video/audio...) a separate unidirectional SVC tree was established with root on the RP router and the branch on the ATM switch in direction to the leaf routers. This tree was used for transmitting in downstream direction from the RP to the leaf routers. The setup of the leaves of these trees was triggered by the workstations that joined the multicast session (WS1 and WS2).

The path used for the upstream data streams from the leaves to the RP has not been fully analysed yet, but first result shows that only a downstream unidirectional SVC tree from the RP router was established. There was no upstream path from the receiver (the leaves) to the sender (the root).

### QoS on SVCs

Using the basic configuration of PIM SM results in the setup of SVCs with UBR service in the ATM switches. Also on the p2mp SVC tree only UBR service is available.

## Protocol Measurement Results



**Figure:** Setup and Release of a multicast group in a ATM-based PIM-SM environment

The protocol measurements were done during testphase 2 using the network infrastructure shown in the above figure. In the above diagram a detailed view of the protocols involved (IGMP, PIM-SM, ATM-UNI) is presented. The diagram is separated into the four steps which are described in more detail in the following

text:

- PIM-SM protocol start.
- Joining of group members.
- Creation of p2mp SVC tree, mapping IP multicast source tree on it, forwarding video traffic.
- Leaving of group members and release of p2mp SVC tree.

### **Step 1: PIM-SM start, establishing of p2p SVCs**

At start time of the PIM-SM protocol e.g. after booting the PIM-configured routers the RP router establishes bidirectional p2p SVCs to each leaf router. These SVCs are used for the PIM protocol message.

### **Step2: Joining of group members**

After announcing a video session via SDR from WS1 this session was visible on the session directories of the other endsystems PC, WS2 and WS3. The signal flow of this session announcement is not shown. The signalling diagram part in step 2 shows the protocol flow that happens when the endsystems join the announced video session.

The join of a endsystem to a multicast group G is fetched by the routers via cyclic IGMP query messages. The endsystems answer with IGMP reports including the multicast group address G they wish to join. Triggered by the IGMP join reports the leaf routers (cisco7000 and cisco7500) are sending PIM-JOIN messages to the RP router (cisco7200). The RP router as well as the leaf routers generates the (\*,G) mroute entry where they register the endsystem and the outgoing interface through which they will forward the multicast traffic for Group G. The PC which acts later as video sender indicates his group membership to the RP router via IGMP and a DVMRP router (WS1).

### **Step3: Creating p2mp SVC tree**

Until this point all the PIM messages were transported on p2p SVCs between the routers. For the multicast transport of the video data from the PC to the leaf endsystems (WS1 and WS2) a p2mp SVC tree is setup. This setup is triggered by the first multicast data packets sent to the group G. These first packets are encapsulated in PIM Register messages and send to the RP-Router, where a (S, G) entry is created for each sender in the group G. After the establishment of the p2mp SVC tree the video data packets for group G are now sent from the PC through the p2mp SVC tree to WS2 and WS3.

The following PIM protocol dump on the RP router shows this process between RP and leaf-1 router in more detail.

#### **RP router**

```
*Mar 31 13:25:47.925: PIM: Received Join/Prune on ATM1/0 from
193.246.0.55, to us
```

```

*Mar 31 13:25:47.925: PIM: Join-list: (*, 224.2.130.69) RP
193.246.0.53
*Mar 31 13:25:47.929: PIM: Check RP 193.246.0.53 into the (*,
224.2.130.69) entry, RP-bit set, S-bit set
*Mar 31 13:25:47.929: PIM: Add ATM1/0/193.246.0.55 to (*,
224.2.130.69), Forward state
*Mar 31 13:25:47.929: PIM-ATM: Send SETUP on ATM1/0 for
224.2.130.69/193.246.0.55
*Mar 31 13:25:48.129: PIM: Forward decapsulated data packet for
224.2.130.69 on ATM1/0
*Mar 31 13:25:48.129: PIM: Send Join on ATM1/0 to 193.246.0.55 for
(193.196.152.3/32, 224.2.130.69)
*Mar 31 13:25:48.145: PIM-ATM: Received CONNECT on ATM1/0 for
224.2.130.69, vcd 9
*Mar 31 13:25:48.501: PIM: Building Join/Prune message for
224.2.130.69
*Mar 31 13:25:50.101: PIM: Forward decapsulated data packet for
224.2.130.69 on ATM1/0
*Mar 31 13:26:48.565: PIM: Send Register-Stop to 193.246.0.55 for
193.196.152.3, group 224.2.130.69
---
193.196.152.3 = WS2
193.246.0.55 = leaf-2 router ATM2/0 IF (cisco7000)
193.246.0.53 = RP router ATM1/0 IF (cisco7200)
224.2.130.69 = group address of video session
vcd9 = virtual channel descriptor for p2mp tree

```

After receiving the PIM-Join from the leaf-1 router on the p2p SVC the RP-router adds its p2p SVC endpoint to the outgoing interface list of the (\*,G) mroute entry. Triggered by receiving of the first multicast data packets from leaf-1 router (which is encapsulated in a PIM Register message) the RP router initiates the setup (sending PIM-ATM Setup) of a p2mp SVC to the leaf-1 router. Also it forwards the decapsulated data packets to the multicast group. After establishing the p2mp SVC (receiving PIM-ATM Connect) the RP sends a PIM Register-Stop message to leaf-1 router where WS2 is connected. Now the video packets origination from the PC are forwarded through the p2mp SVC. At this stage the first SVC tree leaf between RP router and leaf-1 router is established.

Adding the second p2mp SVC leaf to the leaf-2 router happens in the same manner and is not shown in the above protocol dump. It is initiated by receiving a PIM register message which transports encapsulated first multicast data packets from WS3. The RP router initiates the SVC leaf setup by sending ATM-UNI add-party to the directly connected ATM-Switch LS1010-1 which results in establishing of a second p2mp SVC leave between RP and leaf-2 router.

At the end of step 3 a video session was established on group address 224.2.130.69. and the PC sends a video stream to that group, which is received as WS2 and WS3. The copying of the video packets happens at the p2mp SVC branch on the ATM-Switch LS1010-1.



The relevant multicast entries on the RP and leaf-1 router during this video session are shown in the following text.

### **RP router (cisco7200)**

IP Multicast Routing Table

Flags: D - Dense, S - Sparse, C - Connected, L - Local, P - Pruned  
R - RP-bit set, F - Register flag, T - SPT-bit set, J - Join SPT

Timers: Uptime/Expires

Interface state: Interface, Next-Hop or VCD, State/Mode

(\* , 224.2.130.69), 04:15:28/00:02:59, RP 193.246.0.53, flags: SJC

Incoming interface: Null, RPF nbr 0.0.0.0

Outgoing interface list:

ATM1/0, VCD 9, Forward/Sparse, 04:15:08/00:02:36

Tunnel0, Forward/Dvmrp, 04:15:16/00:00:00

(193.196.152.3, 224.2.130.69), 04:15:23/00:02:43, flags: CT

Incoming interface: ATM1/0, RPF nbr 193.246.0.55

Outgoing interface list:

ATM1/0, VCD 9, Forward/Sparse, 00:00:23/00:02:36

(193.196.152.97, 224.2.130.69), 04:14:41/00:02:59, flags: CT

Incoming interface: Tunnel0, RPF nbr 193.196.152.102, Dvmrp

Outgoing interface list:

ATM1/0, VCD 9, Forward/Sparse, 04:14:41/00:02:36

(193.196.152.98, 224.2.130.69), 00:00:15/00:02:44, flags: CT

Incoming interface: Tunnel0, RPF nbr 193.196.152.102, Dvmrp

Outgoing interface list:

ATM1/0, VCD 9, Forward/Sparse, 00:00:15/00:02:44

(\* , 224.2.127.254), 04:15:41/00:00:00, RP 193.246.0.53, flags: SJCL

Incoming interface: Null, RPF nbr 0.0.0.0

Outgoing interface list:

Tunnel0, Forward/Dvmrp, 04:15:25/00:00:00

ATM1/0, VCD 7, Forward/Sparse, 04:15:39/00:02:56

### **Leaf-1 router (cisco7000)**

IP Multicast Routing Table

Flags: D - Dense, S - Sparse, C - Connected, L - Local, P - Pruned  
R - RP-bit set, F - Register flag, T - SPT-bit set, J - Join SPT

Timers: Uptime/Expires

Interface state: Interface, Next-Hop, State/Mode

(\* , 224.2.130.69), 05:09:41/00:02:59, RP 193.246.0.53, flags: SJCF

Incoming interface: ATM2/0, RPF nbr 193.246.0.53

Outgoing interface list:

Ethernet0/1, Forward/Dense, 05:09:41/00:00:00

(193.196.152.3/32, 224.2.130.69), 04:41:00/00:02:44, flags: PCFT

Incoming interface: Ethernet0/1, RPF nbr 0.0.0.0

Outgoing interface list: Null

(193.196.152.97/32, 224.2.130.69), 05:09:41/00:02:59, flags: CJT

Incoming interface: ATM2/0, RPF nbr 193.246.0.53

Outgoing interface list:

Ethernet0/1, Forward/Dense, 05:09:41/00:00:00

(\* , 224.2.127.254), 4d04h/00:00:00, RP 193.246.0.53, flags: SJCL

Incoming interface: ATM2/0, RPF nbr 193.246.0.53

Outgoing interface list:

Ethernet0/1, Forward/Dense, 08:45:07/00:00:00

---

193.196.152.3 = WS2

193.196.152.97 = PC

193.196.152.98 = WS1 Ethernet IF

193.246.0.53 = RP router ATM1/0 IF (cisco7200)

193.246.0.55 = leaf-2 router ATM2/0 IF (cisco7000)

224.2.127.254 = default group address for session announcements

224.2.130.69 = group address of announced video session

VCD 7 = channel descriptor for p2mp SVC tree in default session announcement group

VCD 9 = channel descriptor for p2mp SVC tree in announced video session

The above relevant multicast routing table shows the (\*,G) and (S,G) entries of the video session as well as the (\*,G) entry of the default group for session announcements. The video session group (224.2.130.69) traffic is forwarded on the p2mp SVC with descriptor VCD #9. The session announcements are forwarded on VCD #7.

The ATM SVC state at the end of step 3 is shown in the following router output. On the RP router for each group a p2mp SVC tree is established. The SVC leaves are terminated on each leaf router. The number of leaves on the SVC trees are indicated in the row 'Type'. For the video session multicast traffic the SVC tree with descriptor vcd #9 was used, it is of type MSVC-2 indicating a p2mp SVC with two leaves. A second p2mp SVC tree is established for the session announcement group (VCD

#7) On all SVCs a UBR traffic profile is sent.

### RP router (cisco7200)

```
de-p2mp-router#sh atm vc
          VCD /
Interface Name VPI VCI Type Encaps      Peak Avg/Min  Burst
          Kbps Kbps   Cells  Sts
1/0        5     0 5    PVC SAAL      155000 155000
1/0        16    0 16   PVC ILMI    155000 155000
1/0        2     0 37   SVC SNAP    155000 155000
1/0        4     0 41   SVC SNAP    155000 155000
1/0        6     0 42  MSVC-2 SNAP    155000 155000
1/0        7     0 43  MSVC-2 SNAP    155000 155000
```

Each SVC has different usage depending on the traffic load from the senders, as the following text shows. The highest packet rate is on VCD #9, which is the rate of the video traffic forwarded on the SVC tree.

```
de-p2mp-router#sh ip pim vc
IP Multicast ATM VC Status
ATM1/0 VC count is 7, max is 200
Group          VCD Interface      Leaf Count Rate
224.2.127.254  7   ATM1/0           2         45 pps
224.2.130.69   9   ATM1/0           2        5668 pps
```

### Leaf-1 router (cisco7000)

The p2mp-SVC endpoints are also visible on the leaf-1 router:

```
de-leaf-router#sh atm vc
          VCD /
Interface Name VPI VCI Type Encaps      Peak Avg/Min  Burst
          Kbps Kbps   Cells  Sts
2/0        1888 0   289 SVC  AAL5-SNAP  155000 155000 192  ACTIVE
2/0        1889 0   290 MSVC AAL5-SNAP  155000 155000 192  ACTIVE
2/0        1890 0   301 MSVC AAL5-SNAP  155000 155000 192  ACTIVE
```

## **Step 4: Leaving of group members and release of p2mp SVC**

The leaving of a group member (WS2) is signalled to the directly connected router (leaf-1) via IGMP leave message. The leaf-1 router deletes the (\*,G) mroute entry of the video session group and sends a PIM Prune message to the RP router. After receiving this PIM Prune message the RP router deletes the (S,G) entry, and initiates the release of the p2mp SVC leaf to the leaf-1 router. The following text shows this process in the leaf-1 router and the RP router.

### Leaf-1 router (cisco7000)

```

Mar 31 13:32:28.323: IGMP: Received Leave from 193.196.152.3
(Ethernet0/1) for 224.2.130.69
Mar 31 13:32:28.327: IGMP: Send v2 Query on Ethernet0/1 to
224.2.130.69
Mar 31 13:32:30.039: IGMP: Send v2 Query on Ethernet0/1 to
224.2.130.69
Mar 31 13:32:32.051: IGMP: Deleting 224.2.130.69 on Ethernet0/1
Mar 31 13:32:32.051: PIM: Send Prune on ATM2/0 to 193.246.0.53 for
(193.246.0.53/32, 224.2.130.69), RP-bit
Mar 31 13:32:48.079: PIM: Building Join/Prune message for
224.2.130.69
Mar 31 13:33:48.272: PIM: Building Join/Prune message for
224.2.130.69
Mar 31 13:34:48.376: PIM: Building Join/Prune message for
224.2.130.69
Mar 31 13:35:02.388: PIM: RP 193.246.0.53 for group 224.2.130.69
went down

```

### **RP router (cisco7200)**

```

*Mar 31 13:32:28.749: PIM: Prune-list: (*, 224.2.130.69) RP
193.246.0.53
*Mar 31 13:32:28.753: PIM-ATM: Send RELEASE on ATM1/0 for
224.2.130.69, vcd 9
[.]
*Mar 31 13:32:28.753: PIM: Prune ATM1/0/224.0.0.2 from (0.0.0.0/32,
224.2.130.69, vcd 9
*Mar 31 13:32:28.753: PIM-ATM: Received RELEASE-COMPLETE on ATM1/0
for 224.2.130.69

```

### **Open issues for further investigation**

The described findings and measurements show the p2mp capabilities and basic behaviour of the PIM-SM protocol using p2mp SVCs. There are a lot of topics left for further investigation:

- SVC trees with QoS
- bidirectional SVC trees
- OoS between the endsystems using RSVP
- Native ATM between the endsystems using ATM multicast mechanisms in the LAN (e.g. MARS)
- Verification of the RP auto-discovery protocol

### **Bibliography and References**

- [1] RFC 2117: Protocol Independent Multicast, Sparse Mode
- [2] Cisco IP Multicast EFT & Beta Information  
<ftp://ftp-eng.cisco.com/ipmulticast/html/ipmulticast.html>
- [3] Cisco documentation "Configuring IP Routing Protocols"  
[http://www.cisco.com/univercd/cc/td/doc/product/software/ios112/112cg\\_cr/5cbook/5ciprout.htm](http://www.cisco.com/univercd/cc/td/doc/product/software/ios112/112cg_cr/5cbook/5ciprout.htm)
- [4] Cisco documentation "IP Multicast over ATM Point-to-Multipoint Virtual Circuits"  
<http://www.cisco.com/univercd/cc/td/doc/product/software/ios112/p2mpvc.htm>

## 4.6 ATM Signalling

### 4.6.1 Experiment leader

Christoph Graf, DANTE, Cambridge, UK (since 02/98: Sun Microsystems, CH)

### 4.6.2 Summary of results

As JAMES could still not offer support for native ATM signalling, our tests were again based on PVPC-based tunnelling of ATM signalling information across the JAMES network.

No endsystem support for the latest UNI version was available to us during our tests. We therefore continued to use UNI3.1, instead of UNI4.0, as planned. Furthermore, as the IP stacks in our endsystems support SVCs based on UBR only, we could not test other traffic classes as planned and restricted ourselves to testing UBR. The setup was based on phase 1 experiment »SVC Tunneling through PVPCs«, but where possible, all equipment was upgraded to the latest available firmware/OS releases. With respect to the results gathered in phase 1 testing (cf. D11.3, par. 4.2), a significant increase in SVC setup reliability could be observed. Deployment of signalled ATM Vcs across ATM WAN links in non-mission critical applications could now be envisaged. Critical for further deployment of ATM signalling is reduction of the configuration overhead by successful deployment of distributed, dynamic ATM address resolution and dynamic ATM routing protocols.

### 4.6.3 Participants to the experiment

- ACONET (AT)
  - Gerald Hanusch, Universitaet Linz
  - Guenther Schmittner, Universitaet Linz
- DFN (DE)
  - Robert Stoy, RUS
- INFN (IT)
  - Mauro Campanella, INFN
  - Tiziana Ferrari, INFN/CNAF
  - Simone Maggi, INFN
  - Stefania Alborghetti, INFN
- RCCN (PT)
  - José Vilela, RCCN
- REDIRIS (ES)
  - Celestino Tomas, REDIRIS
- SWITCH (CH)
  - Simon Leinen, SWITCH
- UKERNA (UK)

- Christoph Graf, DANTE
- ULB (BE)
  - Ramin Najmabadi, ULB
- UNINETT (NO)
  - Olaf Kvittem, UNINETT
  - Vegard Engen, BDC, Bergen, Norway (formerly: UNINETT)

#### 4.6.4 Dates and phases

##### Phase one: Preparatory works

Date: 06/97 - 07/97

##### Phase two: Network implementation

Date: 08/97 - 09/97

##### Phase three: Experimentation

Date: 10/97 - 12/97

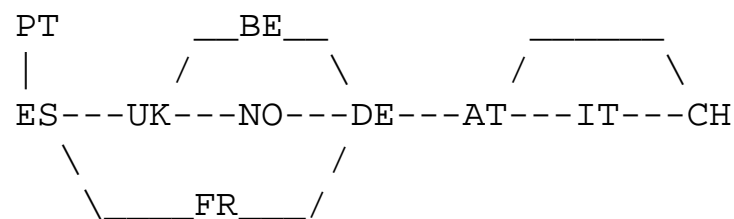
#### 4.6.5 Network infrastructure

None in phase one.

The second and third phase require VPs (CBR or VBR) of 2 Mbps to interconnect the participants. All interconnectivity was based on the »Overlay Network«.

#### 4.6.6 Results and findings

##### Map of SVC connected sites in phase 2 and 3



##### Set-up of ATM equipment in all sites in phase 2 and 3

The physical setup was taken from phase 1 experiment 4.2 (SVC Tunneling through PVPCs) and can be found in deliverable D11.3.

## Phase 1: Preparatory works

For none of our ATM end systems were UNI4.x capable firmware releases available. However, intermediate bugfix releases were available in most cases. It was decided to upgrade all endsystems to the latest available release.

While recent Cisco IOS releases support UNI4.0, only support for UNI3.1 was available for Fore switches. It was decided to upgrade the overlay network to use UNI4.0 on all links terminated by Cisco switches on both sides and to use UNI3.1 (UNI3.0 towards BE) on all remaining links.

## Phase 2: Network implementation

The network as described in D11.3 par. 5.2 was set up.

## Phase 3: Experimentation

The next paragraphs summarise the details about the phase three experiments.

### SVC reliability

The SVC setup reliability was measured in the experiment 5.2 in phase 1 and the result was quite disappointing (see D11.3, par 5.2). The same setup was used in this experiment to verify whether the firmware/OS upgrade of all ATM equipment could yield a significantly better result.

- The tool used for testing connectivity and measuring set-up times is the standard unix tool "ping" with default packet size of 64 bytes ICMP payload.
- Connectivity was considered established, if at least 2 out of 5 ping packets were returned from the remote host within 10 seconds from the last packet submission.
- If not already established, the first packet of a ping sequence opens prior to its submission a SVC to the remote host, while subsequent packets will make use of this connection without the need to establish a SVC. Thus, the delay difference in the round trip time (RTT) between the slowest and the fastest packet of such a ping sequence is a good estimate of the SVC set-up time and the results below use this calculation method.
- Test duration was about 2hrs, average sample size for all results is about 5-7 values.
- The tests were carried out from systems in CH, ES, FR, IT, NO and UK.
- The results given below are median values.

local area connection	remote connection	same host connection
--------------------------	----------------------	-------------------------



## Colour legend

193.246.0. xx	CH (.82)	ES (.102)	FR (.115)	IT (.129)	NO (.184)	UK (.225)
From:						
To:						
AT (.22)	55		x	448	101	178
BE (.32)	221	122	x	281	115	68
DE (.54)	80	80	x	500	46	124
DE (.55)	93	93	x	374	59	139
PT (.73)	232	34	x		129	78
PT (.74)	240	33	x		129	80
PT (.75)	234				120	84
CH (.81)	17	152	x	223	120	199
CH (.82)		152	x	82	117	198
CH (.83)	10	148	x	165		196
ES (.100)	215	17	x			62
ES (.101)	223	19	x			68
ES (.102)	220	1	x			64
FR (.115)	116	44		173	81	89
FR (.116)	120	49	x		86	99
IT (.129)	473	284	x	1	410	385
IT (.132)	81	480	x	36	387	341
IT (.134)	96	789	x	2	469	238
NO (.177)	133	119				73
NO (.181)	120	108		402		63
NO (.184)	116	109	x	221		59
UK (.225)	163	64	x		64	
UK (.226)	169	68	x	257	60	17

Table 2.1: SVC Set-up Times in ms (median values, sample size 5-7)

193.246.0. xx  From: To:	CH (.82)	ES (.102)	FR (.115)	IT (.129)	NO (.184)	UK (.225)
AT (.22)	22		64	54	67	113
BE (.32)	128	40		202	42	11
DE (.54)	57	50	32	126	32	79
DE (.55)	57	50	32	135	32	79
PT (.73)	156	10	128		70	39
PT (.74)	157	10	128		71	39
PT (.75)	156					39
CH (.81)	0	105	84	51	87	134
CH (.82)		104	84	68	87	133
CH (.83)	1	105	84	48		134
ES (.100)	148	2	120			31
ES (.101)	148	1	120			30
ES (.102)	147	0	120			30
FR (.115)	85	20	0	96	61	49
FR (.116)	85	20	1		61	49
IT (.129)	54	121	324	0	177	148
IT (.132)	51	174	156	1	160	199
IT (.134)	48	163	140	1	324	199
NO (.177)	94	69				39
NO (.181)	87	62		124		32
NO (.184)	87	62	60	113		32
UK (.225)	133	30	48		32	
UK (.226)	119	31	92	127	34	2

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

193.246.0. xx	CH (.82)	ES (.102)	FR (.115)	IT (.129)	NO (.184)	UK (.225)
From:						
To:						
AT (.22)	1		1	1	1	1
BE (.32)	1	1		1	1	1
DE (.54)	1	1	1	1	1	1
DE (.55)	1	1	1	1	1	1
PT (.73)	1	1	1		1	1
PT (.74)	1	1	1		1	1
PT (.75)	1					1
CH (.81)	1	1	1	1	1	1
CH (.82)		1	1	0.8	1	1
CH (.83)	1	1	1	1		1
ES (.100)	1	1	1			1
ES (.101)	1	1	1			1
ES (.102)	1	1	1	0.4		1
FR (.115)	1	1		1	1	1
FR (.116)	1	1	1		1	1
IT (.129)	1	0.8	1	1	0.86	1
IT (.132)	1	1	1	1	1	1
IT (.134)	0.8	1	1	1	0.71	1
NO (.177)	1	1				1
NO (.181)	1	1		1		1
NO (.184)	1	1	1	1		1
UK (.225)	1	1	1		1	
UK (.226)	1	1	1	1	1	1

Table 2.3: Probability that a SVC can be established

## Comments

- Due to processing overhead of signalling messages inside the switches, the SVC set-up times are always substantially above the theoretical lower bound of one RTT. This can be a major drawback in those applications, where fast response times are critical (e.g. name service queries,

DB queries).

- The problems experienced on the connections to and from Italy were caused by cell loss on an ATM VP, which had to be shared between different applications. Traffic shaping could not be performed on the aggregated bandwidth, but only on each source separately, thus causing cell loss.
- The IP-NSAP mapping on some end systems was incomplete as was the ATM routing table on some switches, resulting in the number of empty entries in the tables above. They were not caused by signalling problems
- Testing in FR was performed on a Cisco router, which does unfortunately drop the first ping packet (when sent from the command line interface and when forcing an SVC to be established). It was therefore not possible to gather information about the SVC setup times from FR.

### 5.6.7 Major observations during the tests

D11.3 par. 5.2 describes the results gained with the same setup with quite disappointing results. After upgrading all ATM equipment with the latest firmware/OS, far better results were achieved. This proves, that the industry is taking ATM seriously and putting effort into support of ATM. Unexplained reboots of equipment or other irregularities, were no longer observed.

### 5.6.8 Relevance for service and outlined migration to service

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

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

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

Our tests show, that a signalling infrastructure based on the equipment at hand right now, is becoming feasible. This is one small step towards the ultimate goal of ATM: to offer a global infrastructure offering QoS-based interconnects between applications.

### 5.6.9 Test-related problems and general comments

- The JAMES procedures and the configuration overhead involved in setting up new VPs between our sites proved to be a too complicated and lengthy process to be able to order VPs at short notice as needed. Therefore, wherever possible, the "overlay network" was used for our tests.
- In order to eliminate any impact of ATM address resolution to the SVC setup time, static IP-NSAP address mapping was used (instead of e.g. ATMARP). Since PNNI did not work properly on FORE switches during our tests, IISP was used on all links involving FORE switches. As a consequence, all end systems had to be configured with an up to date mapping table containing all participating end systems and all switches had to be loaded with a complete set of all ATM routing entries. With new systems being added and ATM addresses being updated, this was not the case all of the time and resulted in missing entries in the reachability matrix.

### 5.6.10 Further studies

- The only application tested so far was the ATM/AAL5/IP stack. Other applications should be considered as well.
- Our switches support currently only switching of UBR Vcs. Other traffic classes should be considered too as they become available.
- JAMES did so far not support native signalling support. It is not well understood whether and how ATM service provider will offer signalling support to customers. Careful monitoring of future developments is necessary.

### 5.6.11 Annex

#### Static IP to NSAP mapping and NSAP prefix table

```
# Mapping between IP and NSAP addresses for SVC testing over JAMES
# and NSAP prefixes used on involved switches
#=====
# last update: 29/01/98 CG
#
# IP address NSAP address
#
# ACONET (AT)
prefix: 39.040F.5404.0101.XXXX.XXXX.XXXX.XXXX.XXXX.XXXX.XX
193.246.0.18 39.040F.5404.0101.0001.0001.0001.0020.4804.F6DF.01
193.246.0.20 39.040F.5404.0101.0001.9999.0001.0020.EA00.0B22.00
```

```
193.246.0.21 39.040F.5404.0101.0001.9999.0001.1111.1111.1111.50
193.246.0.22 39.040F.5404.0101.0001.9999.0001.9999.9999.9901.50
#
# ULB/STC (BE)
prefix: 39.056F.XXXX.XXXX.XXXX.XXXX.XXXX.XXXX.XXXX.XXXX.XX
193.246.0.32 39.056F.0000.0000.0000.0000.0000.0001.9324.6032.01
#
# Belgacom (BE)
prefix: 47.0005.80FF.E100.0000.F215.100F.XXXX.XXXX.XXXX.XX
193.246.0.40 47.0005.80FF.E100.0000.F215.100F.0020.4815.100F.00
#
# DFN/RUS (DE)
prefix: 39.276F.3100.0110.0000.0001.XXXX.XXXX.XXXX.XXXX.XX
193.246.0.54 39.276F.3100.0110.0000.0001.0003.0020.4806.0989.01
193.246.0.55 39.276F.3100.0110.0000.0001.0003.1111.1111.1102.04
#
# RENATER (FR)
prefix: 39.250F.0000.002D.XXXX.XXXX.XXXX.XXXX.XXXX.XXXX.XX
193.246.0.115 39.250F.0000.002D.0101.0101.0101.1932.4600.0115.01
193.246.0.116 39.250F.0000.002D.0101.0101.0101.1932.4600.0116.01
#
# RCCN (PT)
prefix: 39.620f.0000.0000.0000.0000.0000.XXXX.XXXX.XXXX.XX
193.246.0.73 39.620f.0000.0000.0000.0000.0000.0020.4806.84b9.01
193.246.0.74 39.620f.0000.0000.0000.0000.0000.0020.481a.3714.01
193.246.0.75 39.620f.0000.0000.0000.0000.0000.0000.0000.0013.00
193.246.0.76 39.620F.0000.0000.0000.0000.0000.0010.11BB.D701.01
#
# SWITCH (CH) (see also DNS zone: tf-ten.switch.ch)
prefix: 39.756F.1111.1111.7001.0001.1002.XXXX.XXXX.XXXX.XX
193.246.0.81 39.756F.1111.1111.7001.0001.1002.1932.4600.0081.01
193.246.0.82 39.756F.1111.1111.7001.0001.1002.1932.4600.0082.01
193.246.0.83 39.756F.1111.1111.7001.0001.1002.1932.4600.0083.01
#
# REDIRIS (ES)
prefix: 39.724F.10.010001.0001.0001.0001.XXXX.XXXX.XXXX.XX
193.246.0.101 39.724F.10.010001.0001.0001.0001.1932.4600.0101.00
193.246.0.102 39.724F.10.010001.0001.0001.0001.0020.4806.225B.00
193.246.0.103 39.724F.10.010001.0001.0001.0001.1932.4600.0103.02
#
# INFN (IT)
prefix: 39.380F.XXXX.XXXX.XXXX.XXXX.XXXX.XXXX.XXXX.XXXX.XX
193.246.0.129 39.380F.0000.0000.0000.0000.0000.0019.3246.0129.01
```

```

# 193.246.0.130 39.380F.0000.0000.0000.0000.0000.0019.3246.0130.01
# 193.246.0.131 39.380F.0000.0000.0000.0000.0000.0019.3246.0131.01
193.246.0.132 39.380F.0000.0000.0000.0000.0000.0019.3246.0132.01
# 193.246.0.133 39.380F.0000.0000.0000.0000.0000.0019.3246.0133.01
193.246.0.134 39.380F.0000.0000.0000.0000.0000.0020.4815.15A9.01
# 193.246.0.135 39.380F.0000.0000.0000.0000.0000.0019.3246.0135.01
#
# CNAF (IT)
prefix: 39.380f.1001.0001.0000.0001.XXXX.XXXX.XXXX.XXXX.XX
193.246.0.138 39.380F.1001.0001.0000.0001.0000.0019.3246.0138.01
193.246.0.139 39.380F.1001.0001.0000.0001.0000.0019.3246.0139.01
193.246.0.140 39.380F.1001.0001.0000.0001.0000.0019.3246.0140.01
#
# RESTENA (LU)
prefix: 39.442F.XXXX.XXXX.XXXX.XXXX.XXXX.XXXX.XXXX.XXXX.XX
193.246.0.144 39.442F.0000.0000.0000.0000.0001.0020.481A.1D5B.00
193.246.0.145 39.442F.0000.0000.0000.0000.0001.0020.4806.221E.00
#
# UNINETTT (NO)
prefix: 47.0023.0100.0005.XXXX.XXXX.XXXX.XXXX.XXXX.XXXX.XX
193.246.0.177 47.0023.0100.0005.2000.0001.0101.1034.1034.1034.01
193.246.0.178 47.0023.0100.0005.2000.0101.0120.0800.093d.0641.00
193.246.0.181 47.0023.0100.0005.4000.0001.0101.2000.0001.0001.70
193.246.0.184 47.0023.0100.0005.4000.0001.0101.0800.093d.063c.01
#
# UKERNA (UK)
prefix: 39.826F.1107.2500.10XX.XXXX.XXXX.XXXX.XXXX.XXXX.XX
193.246.0.225 39.826f.1107.2500.1000.0000.0000.0020.4806.1ff1.00
193.246.0.226 39.826f.1107.2500.1000.0000.0000.0020.481a.2e52.01

```

## 5.6.12 References

1. ATM Forum, "ATM User-Network Interface Specification Version 3.0", 1993
2. ATM Forum, "ATM User-Network Interface Specification Version 3.1", 1994
3. ATM Forum, "ATM User-Network Interface Specification Version 4.0", 1996
4. J. Heinanen, "Multiprotocol Encapsulation over ATM Adaptation Layer 5", RFC 1483, Telecom Finland, July 1993
5. M. Laubach, "Classical IP and ARP over ATM", RFC 1577, Hewlett-Packard Laboratories, January 1994
6. M. Perez et al., "ATM Signaling Support for IP over ATM", RFC 1755, USC/Information Sciences Institute, February 1995

## 4.7. ATM Policy control and accounting

### Experiment leader:

Victor Reijs, SURFnet bv, the Netherlands

### Summary of results

This experiment examined how SVCs in an ATM Network can be controlled on a Policy level. The basic issue is that ATM SVCs claim resource from the network. Therefore, the NRN and their connected institutes need a method to carefully control the user's access to that resource. We call the rules for that access a access policy. The mechanism to enforce those rules is called Policy Control.

In section 1, the need for an SVC admission Policy is examined. Then three points in time relative to the lifecycle of an SVC are identified where Policy Control can be performed. These are

1. Upon SVC Setup Request (section 1.3.1)
2. Interfering With Existing SVCs (section 1.3.2)
3. Checking After The Fact (section 1.3.3)

In section 2 solutions to Policy control are investigated. Ways to implement Policy Control are examined, and four areas of criteria on which to base Policies are identified.

In section 3 the current state of the art in Policy Control is discussed. In this section the IETF Resource ReserVation Protocol (RSVP) and ATM Accounting are discussed.

From this work it can be concluded that having some form of an SVC admission Policy is a requirement for introducing a native ATM SVC service. In this work some criteria on which to base Policy decisions have been identified. Current ATM equipment does not seem to support any form of Policy Control. The IETF is currently doing work on RSVP and RAP, but how to apply this work for ATM SVC Policy Control is currently not defined. ATM accounting might contribute to ATM Policy Control, but implementations of complete standards for ATM accounting are currently unavailable.

The need for an admission Policy in a network that allocates resource to connections to guarantee QoS is not specific to ATM SVCs. An IP/RSVP network also reserves resource to provide QoS, so RSVP will be facing the same type of problems. Therefore, the work done in this experiment will also be applicable to RSVP. Groups working on ATM SVC Policy control therefore have their goal in common with groups working on IP/RSVP and IP Policy Control, and should therefore be in close cooperation with each other.



## Participants:

SURFnet bv, University of Twente and SWITCH

## Results and findings

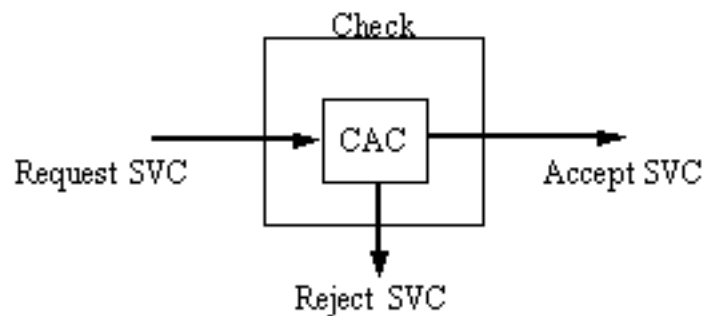
### 1. Description

This experiment examined the consequences of the introduction of an ATM SVC service in a production ATM network environment, in particular the consequences of the ability of the user to claim network resources (by setting up SVCs with a guaranteed QoS). A study of the current State of the Art in this area is included.

#### 1.1 The SVC Service

This section provides a short overview of those aspects of the ATM SVC service that are of importance to be able to discuss ATM SVC Policy Control.

- The SVC service enables the end user to set up ATM connections on demand. At this point a distinction must be made between users and endstations. In this document the term user is used to identify a human being. The term endstation specifies a computer system that is at the end of an ATM link. An endstation can be referenced by its ATM address. Therefore a SVC setup is always initiated by an end-user, and the resulting SVC connects two or more endstations.
- a SVC claims network resources. These resources include bandwidth on the ATM links in the network, available VPI and VCI values, cell buffer space in the ATM switches etc. These resources are all shared and finite.
- When a user requests a new SVC from an ATM network, the network will check if there are enough network resources available to accommodate that SVC. If this is the case, the SVC setup request is accepted, otherwise it is rejected. This function is known in current ATM standards as Call Admission Control (CAC).



*Figure 1: Logical view of the Call Admission Control function of an ATM network*

## 1.2 The need for an SVC Policy

With ATM switches currently on the market it is possible to build an ATM network that provides an SVC service. This SVC service will do Call Admission Control, so a request for a SVC will only succeed if the network has sufficient resources to meet the Quality of Service guarantees for that SVC.

The main topic of this document is that having only a CAC function in the network is not sufficient for a production ATM network. Additional mechanisms are needed to control the acceptance of SVCs, and thus how the scarce network resources are divided among the users. We will refer to these additional mechanisms as the SVC acceptance policy or SVC Policy for short.

An ATM SVC network with only CAC (which can be build using ATM equipment available today) can be regarded as having a simple "first come first served" SVC Policy. All users have equal chances of claiming and using network resources, on a first come first served basis. A user can at any point in time claim all remaining network resources and keep them occupied for an arbitrary long period of time. In a production network, where users actually pay for the network service, such a policy will not be acceptable.

Consider as an example a 155 Mbit/sec university ATM network that serves students as well as staff members. Say the network is completely empty on sunday evening. A student requests a 155Mb CBR SVC. The request is accepted by the CAC, because the network can currently support the connection. The student does not release the connection. On monday morning, no student, staff member or professor is able to set up an SVC, because all resources are still in use by the student. This is clearly an unwanted situation. Other mechanisms are needed to support more detailed and refined policies than the simple "first come first served" policy provided by CAC.

### 1.2.1 SVCs compared to PVCs

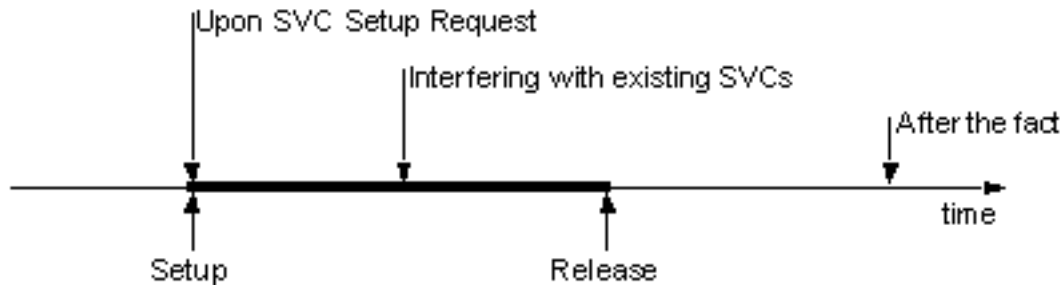
Management (instead of signalling) can also be used to create connections in ATM networks. In ATM terminology such connections are called PVCs. These PVCs claim network resources, just like their SVC counterparts, so in principle the same Policy Control issues as in the SVC case apply.

The key difference between PVCs and SVCs is that PVCs are not actually created by the end-user; they are created by a manager instead. Thus, the decision weather to accept a PVC is made by that manager also. The manager must check if the required resources are available, and if the PVC complies with the PVC acceptance policy. As a result, for PVCs Policy Control is automatically present in the system; the PVC Policy is the set of rules used by the manager to accept or deny a user request for a PVC.

## 1.3 The location of Policy Control in the SVC Lifecycle

An SVC is setup at some point in time, then exists for a period of time, and finally the SVC is released. We will call this the Lifecycle of the SVC. In this lifecycle we identify three different sections where

Policy Control can be applied.



*Figure 2: Policy Control in the SVC lifecycle*

Policy Control can be applied in the SVC lifecycle upon the SVC Setup Request, while the SVC is active, or after the SVC has been released. Each of these will be discussed in a separate section below.

### 1.3.1 Checking Upon SVC Setup Request

When Policy Control is to be performed the moment the network receives a SVC setup request, two things have to be checked. These are the normal CAC check and the Policy Control check. Because the ATM network sets up a SVC on demand with a sub-second setup time, both checks must necessarily be performed in that same timeframe, and therefore in an automated manner.

## Connection Admission Control

Connection Admission Control (CAC) ensures that a new connection is only accepted if the network has sufficient resource to accommodate that particular new connection, without degrading the service level for the already existing connections. Resource of interest in this respect is e.g. bandwidth, cell buffer space in switches, etc. Standards for ATM CAC already exist See [UNI31] and current ATM switches support them.

## Policy Control

The task of Policy Control is to divide the available network resource in a defined manner over the set of network users competing for that resource. Policy Control enforces a connection admission policy. A connection admission policy is a set of rules defining under what circumstances (other than the mere availability of the requested resource) the request for a SVC is to be accepted.

### 1.3.2 Interfering with existing SVCs

Another way to enforce a Policy in a more ad-hoc manner is to allow a manager to interfere with existing SVCs. The manager in this case could be either a human manager or a software management

application. The manager could free up resources by deleting SVCs, thereby making the free resources available for other SVCs. To do this, the manager should be able to

- get an overview of existing SVCs;
- delete a particular (unwanted) SVC;
- prevent a deleted SVC from being re-setup immediately.

For this form of Policy Control the manager would need a set of software tools that provide the functionality listed above.

### 1.3.3 Checking after the fact

The checking after the fact method is an additional solution to the problem of controlling SVCs. It inherently cannot prevent users from claiming all of the available remaining resource; since it works only after a SVC has been setup and destroyed again. Instead, this mechanism ensures that the network users know that their usage of the network is somehow measured. These measurements can then e.g. be used to send bills to users or groups of users.

## 2. Solutions to Policy Control

To be able to enforce a Policy, the ATM network must be extended with a Policy Control function. This chapter provides an architectural framework to define and implement such policies.

### 2.1 An Implementation Architecture for Policy Control

There are three options for the location of the Policy Control check relative in time to the CAC check; either before CAC, after CAC or in parallel with CAC. In general a combination of information on individual users, accounting information and global network status will be needed to be able to make a Policy decision. To avoid storing copies of all this information at each point in the network where Policy decisions must be taken, a network wide Policy Server is introduced. This Policy Server will hold all the information needed to enforce the network SVC acceptance Policy, and switches will contact this server for their information. This is shown in the figure below.

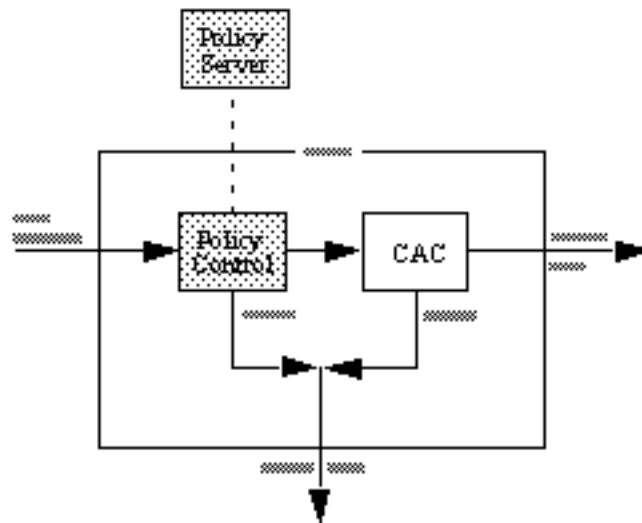


Figure 3: A switch with Policy Control

## 2.2 The Policy Areas

Policy decisions are taken based on a set of criteria. In this section four areas of such criteria are identified. Also the architecture of a possible implementation for Policies using those criteria is presented.

### 2.2.1 Specific Users or Endstations

The decision to accept a request for a new SVC can be based on which particular human user or which particular ATM endsystem is requesting the connection. Ideally, one would like to base policies on human users rather than on ATM endsystems. It is to be expected that implementing Policies based on specific users is more difficult than implementing Policies based on specific ATM endstations. Identifying a specific human user would have to be part of the ATM signalling system.

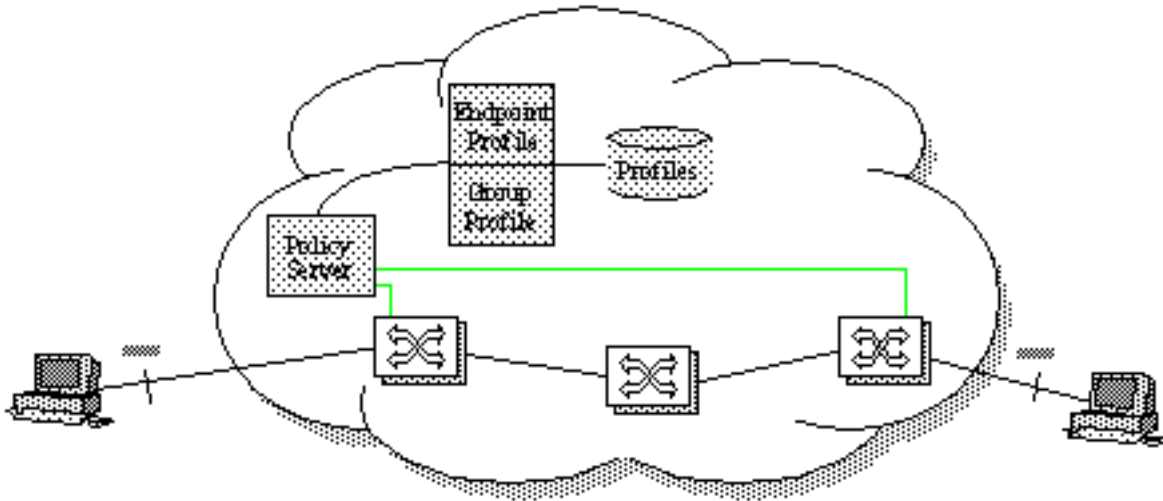
An example of a Policy based on specific users is a policy where users in the group 'students' are allowed connections up to 1 Mbit/sec, and users in the group 'staff members' are allowed connections up to 10 Mbit/sec.

To be able to make a Policy decision based on which particular endpoint or which particular user is requesting the new SVC, the Policy server needs information on the endpoints and users that exist. Each endpoint (or each user) has a certain 'profile' attached to it. This profile contains parameters for the Policy decision, e.g. maximum allowed bandwidth per connection. The ITU-T has defined an extension to the Q.2931 [Q2931] signalling system that enables the generation and transport of end station identifiers. This extension is defined in the Q.2941.1 and Q.2941.2 documents [Q2941.1] [Q2941.2].

A notion of groups is also modeled in the database of endpoints and users. This allow Policies to take into consideration that groups of similar users or endpoints exist, and to check criteria that apply for the

whole group. An example of a policy that uses this when all members of the group 'students' are allowed to use only a total bandwidth of 50 Mbit/sec.

The policy server has access to a database of endpoints and their associated profiles. This is shown in figure 4 below.



*Figure 4: Individual users policy*

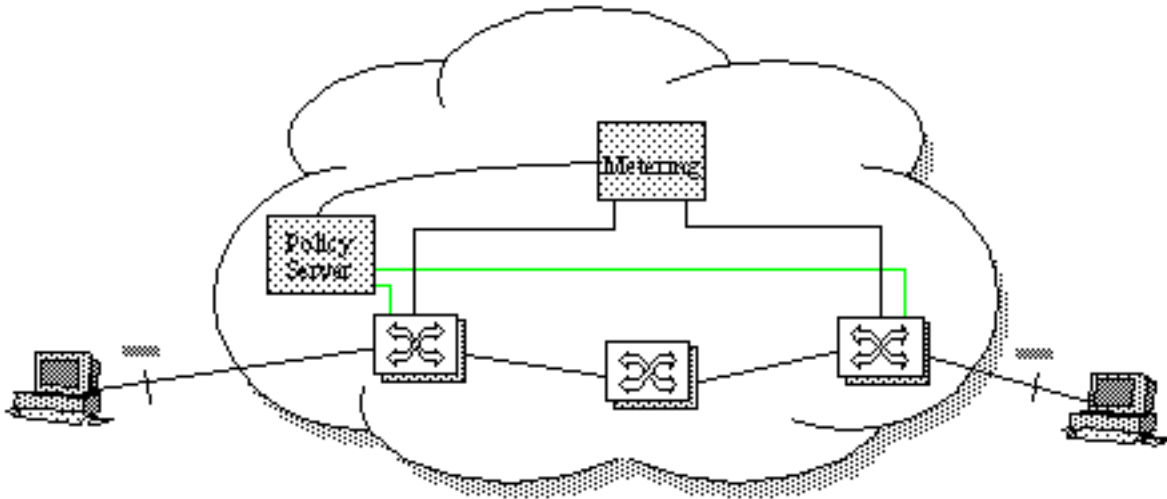
### 2.2.2 Remaining Credits

Policy decisions can also be based on the amount of resource the request endpoint or the destination endpoint has used in the past. As such this criterium for Policy decisions is an extension to the Policies based on specific users or endstations of section 2.2.1.

A new connection can be accepted or rejected based on how much resource this endpoint (or the group the endpoint belongs to) has already used over the last measurement period. As an example an individual endpoint could be allowed to own connections for a total of 100 Mbps-hour per month. This is e.g. a 1 Mbps connection for a period of 100 hours, or a 10 Mbps connection for 10 hours. At the start of a new period the endpoint gets additional credit for setting up SVCs. When the endpoint actually sets up SVCs that credit is debited. For this to work, the usage of network resource by each (group of) endpoint(s) needs to be administrated over time.

Note that credits are not exactly the same as accounting plus billing. With accounting plus billing the user, belonging to the endpoint, must pay for the used resources some time after he has used them. There is however no limit to how much resource a user can use in a single billing period. The combination of accounting and billing is a pshycologically restricting mechanism, where a credit based policy actually prevents additional connections to be set up once the user has used up all his credit.

To implement Policies based on usage credit, the network needs to gather usage information from the network for each endpoint or group of endpoints, as shown in the figure below:

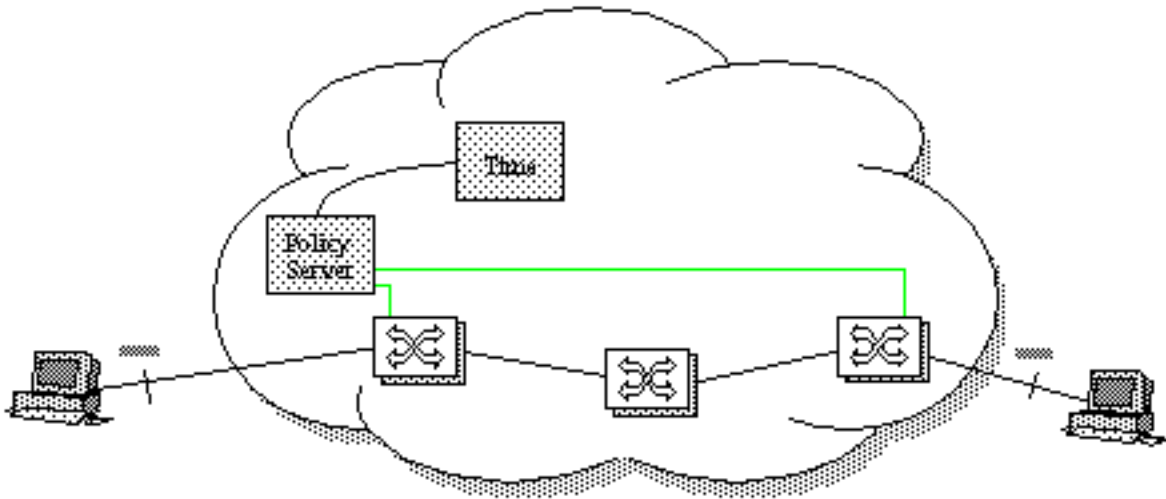


*Figure 5: Remaining credits policy*

### 2.2.3 Current Time

The current time can be used as one of the criteria for the decision to accept or reject the request for a new SVC. An example of a time dependant policy is when student group endpoints are allowed connections of 1 Mbps during business hours, and connections of 2 Mbps outside business hours.

The network Policy server needs access to the current time for time dependant policies, as shown in the figure below.



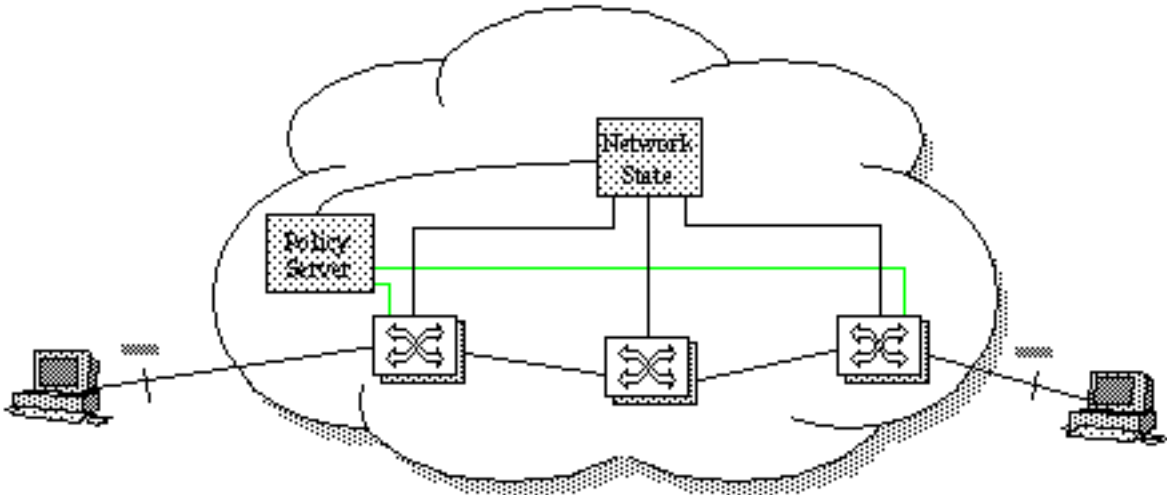
*Figure 6: Current time policy*

Time dependencies of the policy should be implemented in the policy server. A time dependency will always be used in conjunction with some other criterium, like the maximum allowed bandwidth per user in the example.

### 2.2.4 Current network status

The current status of the network can be taken into account when taking Policy decisions. An example of such Policy is when there are two user groups with different priority; low priority endpoints are not allowed to set up SVCs when the network is over 60% loaded, where high priority endpoints are always allowed to set up new SVCs.

To be able to take the current network status into account when taking Policy decisions, the network status must constantly be measured, and the Policy control function must have access to the network status information, as shown in the figure below.



*Figure 7: Current network status policy*

### 3. State of the Art in Policy Control

In this chapter an overview is given of the current state of the art in policy control technology. The first paragraph describes the RSVP world approach.

#### 3.1 RSVP

In the internet world RSVP is used to control the access of endpoints to the network. When the endpoint makes a request for a connection with a certain QoS (in RSVP this is called resource reservation), RSVP carries the request through the network, using the same routing mechanism as the connection will use when it is created. At each visiting node that supports RSVP an attempt is made to allocate the requested resources and to check if the endpoint has enough administrative permissions to make this reservation. If they both succeed, the connection is made, otherwise the endpoint gets a notification that there are not enough resources for his request.

The two modules in the resource allocation mechanism are: admission control and policy control. The admission control module's responsibility is to check if this node has sufficient available resources to supply the requested QoS. This module is similar to the Call Admission Control module in an ATM



network. The policy control module checks if the user has administrative permission to make this reservation.

Reservation protocols, by definition, discriminate between users, by providing some users with better service at the expense of others. Therefore, it is reasonable to expect that these protocols be accompanied by mechanisms for controlling and enforcing access policies. The IETF has already started work for the RSVP protocol in this direction [RAP].

For this experiment the admission control module is of less interest, only the policy control module will be discussed.

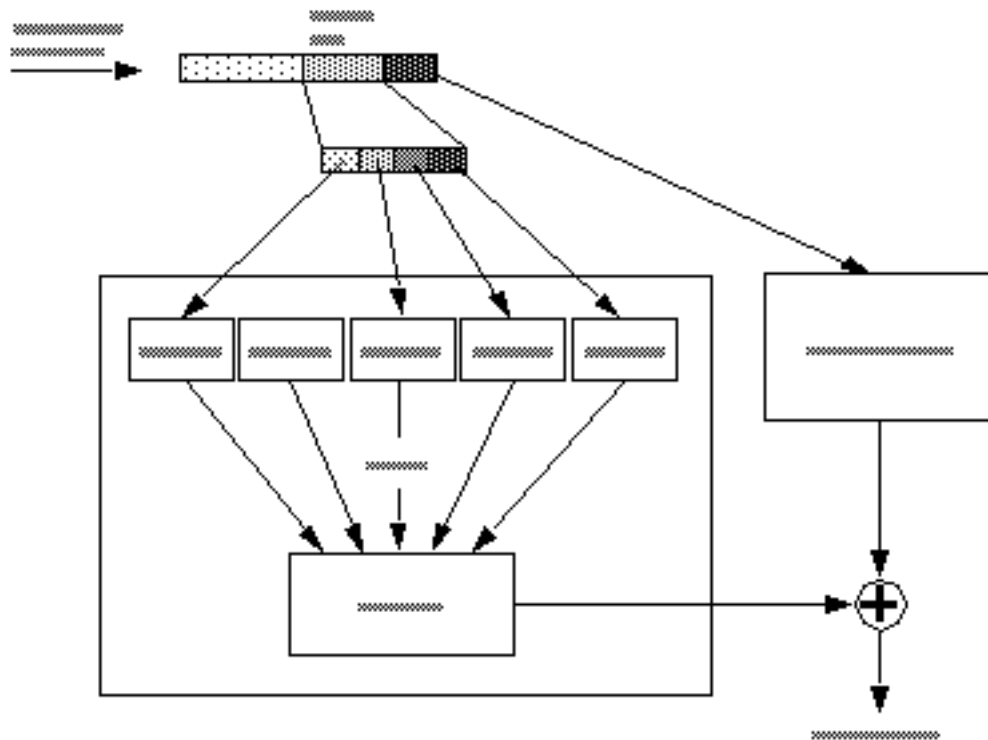
### 3.1.1 Policy Control

The policy control mechanism of RSVP is designed to offer a very flexible mechanism to experiment and develop all kinds of policies. In order to achieve this, the policy space is divided into several (small) policies that can be combined together to offer the desired behaviour. Each policy is responsible for a small and clear policy that handles one aspect of the desired policy. For instance, one of the proposed standard policies is a string credentials check [LPM]. It checks if a certain string, given by the user, is correct/known.

Each policy is identified by its description type, its P-type [LPM]. In the RSVP RESV pdu, that specifies the requested connection, a list of the desired P-types and their attributes is located. There are a number of standard defined policies, with defined P-types. Beside these standard policies it is possible for manufactures to define their own policies. The total policy space is 65535 policies wide.

A RSVP node will implement all, or a subset of all the available policies. Each implemented policy has its own handler. When a RSVP-PDU arrives, it will be given to the admission control module to check if it is technically possible to grant this connection request. If so, it will be given to the policy control module. In here the list of P-types is substracted from the pdu and each P-type is given to its corresponding handler. Each handler will examine the attributes belonging by this policy and will place its vote. The vote can be: accept (this policy is passed), snub (this policy is not passed, but someone else may accept) or veto (this policy is not passed and that is unexceptable). All these votes are combined into a single decision that indicates if this reservation is granted or rejected [OOPS]. The rule used to combine the votes is: *A reservation is approved if there is at least one handler that accepts it, and no other that veto-ed it* [LPM].

If a node has no handler for a certain policy, the complete request will be forwarded to the next RSVP node on the path. This is no problem because RSVP is by design able to work with a mix of RSVP and non-RSVP nodes. Therefore if a node has a certain handler not implemented it looks like if that was a node with no RSVP support.



*Figure 8: RSVP Policy Control*

### 3.2 ATM Accounting

A detailed overview of ATM accounting including background information and an overview of the current state of the art by A. van Dijk [VDIJK] can be found on the web.

Various solutions for ATM accounting are currently under development, by the IETF, the ATM Forum and some ATM Switch vendors.

The IETF [IETF] has two internet drafts available on accounting: "Accounting Information for ATM Networks, November 1996" and "Managed Objects for Controlling the Collection and Storage of Accounting Information for Connection-Oriented Networks, November 1996".

The ATM Forum [ATMF] includes in its B-ICI specification detailed information on the usage of the ATM network, that can be used for accounting purposes.

Some vendors of ATM equipment have accounting built into their ATM products. These vendors include Fore [FORE], Cisco [CISCO], General DataComm [GDC] and Newbridge [NBRIDGE].

### References

[ATMF] The ATM Forum, <http://www.atmforum.com>

[CISCO] Cisco, <http://www.cisco.com>

[FORE] Fore Systems, <http://www.fore.com>

[GDC] General Data Comm, <http://www.gdc.com>

[IETF] The Internet Engineering Task Force, <http://www.ietf.org>

[LPM] Local Policy Modules (LPM): Policy enforcement for Resource Resevation Protocols. Internet-Draft, draft-ietf-rsvp-policy-lpm-01.[ps,txt], Nov. 1996.

[NBRIDGE] Newbridge, <http://www.newbridge.com>

[OOPS] Open Outsourcing Policy Service (OOPS) for RSVP. Internet-Draft, draft-ietf-rsvp-policy-oops-00.[ps,txt], Sept. 1997.

[RAP] R. Yavatkar, D. Pendarakis et al., A Framework for Policy-based Admission Control, IETF Internet-Draft, draft-ietf-rap-framework-00.txt, November 20, 1997.

[UNI31] ATM Forum, User Network Interface Specification version 3.1, july 1996.

[VDIJK] ATM Accounting Management, Thesis of A. van Dijk, <http://www.snp.cs.utwente.nl/~nm/projects/ut-atm/accounting/>

[Q2931] ITU-T Recommendation Q.2931, "B-ISDN DSS 2 UNI Layer 3 specification for basic call/connection control"

[Q2941.1] Draft new ITU-T Recommendation Q.2941.1, "DSS2 Generic Identifier Transport"

[Q2941.2] Draft new ITU-T Recommendation Q.2941.2, "DSS2 Generic Identifier Transport"

## 4.8. ATM traffic management

### Experiment leader

Victor Reijs, SURFnet bv, the Netherlands

### Summary of results

The main goal of this activity is get an idea how existing (UBR, CBR, VBR, etc.) and new (ABR) functionality's will work in the LAN/WAN ATM networks. This work has been divided in two main parts: a) theoretical survey on the different service categories defined by ATM Forum/ITU-T and b) practical work on concatenation of CBR/ABR networks and c) very basic ABR tests.

The concatenation of CBR and ABR networks have been done using proprietary ABR from Digital. The most important results of this experiment is that there are two things to keep in mind when coupling these to different services together:

1. towards the CBR network one needs to do traffic shaping at the boundary ATM switch, otherwise most traffic will be policed away. Traffic shaping at only the end station is not an option in large networks, because one is not able to manage all these end stations with regard to bandwidth in a proper way.
2. because of a possible bottleneck with regard to bandwidth on the CBR WAN network, the end stations on the LAN need to get feedback on these possible bottlenecks. This can be done by ABR.

So the conclusion for these network combinations (CBR/ABR); one needs on the LAN site ABR (preferable VS/VD) to flow control the traffic and one needs to do traffic shaping on the (management) boundary between CBR and ABR network to guard against contract violations.

The most important result from the survey on ABR is that implementations are just emerging on the market! So no real tests have been done with commercial equipment. But beside this lack of extensive practical experience with ABR, the understanding of the options within ABR (EFCI, RR, ER and VS/VD) are much better now. Furthermore we know now that a lot of switches used by NRN's and PNO's will at least transport important cells for ABR (the RM-cells) transparently over the ATM networks.

### Participants

SURFnet bv, NL (V.M.M. Reijs), University of Geneva; Services Informatiques, CH (A. Schindler), University of Stuttgart; National Supercomputing Center, DE (R. Stoy). University of Twente; Centre

for Telematics and Information Technology, NL (P.F. Chimento, E. Meeuwis), University of Utrecht; Department of Physics and Astronomy, NL (H.M.A. Andree, V.J. Giesing, C.T.A.M. de Laat and J. Venema)

## Dates and phases

- understand ABR and other service categories: June 1997 - February 1998
- test concatenation of CBR and proprietary ABR networks: July 1997 - October 1997
- test part of ABR (RM-cells) in WAN: December 1997 - February 1998

## Results

1. [Survey of different ATM service categories](#)
2. [Test concatenation of networks with different ATM categories](#)
3. [Test ABR in LAN and WAN.](#)

### 1. Survey of different ATM service categories

The goal of this deliverable is to explain the latest designs for the various ATM service categories (from the ATM Forum) and to discuss some of the advantages and disadvantages of each. Here we will focus in particular on the ABR and GFR service categories. These are the newest ones and have the closest relationship to data communications (e.g. TCP/IP) over ATM.

### CBR

The CBR service category applies to connections which require cell loss and delay guarantees. The resource (bandwidth) provided to the connection is always available during the connection lifetime, and the source can send at the peak cell rate or below the peak cell rate or not at all.

A CBR connection must specify the parameters found in Table [1](#):

Traffic parameters	How derived
PCR (CLP=0+1)	Derived from the ATM Traffic Descriptor IB or from Network Management
CDVT	Set via Network management procedures and determined on the basis of local customer premises equipment

**Table 1:** CBR parameters

In addition, a CBR connection may specify the QoS parameters: Peak-to-peak cell delay variation, Maximum cell transfer delay and Cell loss ratio.

There is only one conformance definition in [18] for CBR. It specifies that the conformance check is equivalent to a GCRA with parameters  $T_{0+1}$  and  $CDVT$  where  $T_{0+1} = 1/PCR_{0+1}$ . If a cell conforms to the GCRA with these parameters, it is conformant. Since with this conformance definition, no separate parameters for only the  $CLP=0$  stream are specified, tagging is not possible. So, the CLR objectives apply to the combined  $CLP=0+1$  stream. It would be possible to send a conforming stream at PCR which consists only of  $CLP=1$  cells.

## VBR

The VBR service category is intended for a wide range of connections; it includes real-time constrained connections as well as connections that do not need timing constraints. This category is subdivided along just those lines. That is, there is a rt-VBR and nrt-VBR defined. The only difference between them is the sort of QoS parameters that they specify. rt-VBR specifies the same QoS parameters as CBR, i.e. Cell loss ratio, peak-to-peak cell delay variation and maximum cell transfer delay. nrt-VBR specifies only CLR as a QoS parameter. Both types of VBR service support statistical multiplexing. The wording of the specification is such that whether this is done and how it is done is completely unspecified.

The traffic parameters for both VBR service categories are the same, and are shown in Table 2.

Traffic Parameter	How derived
PCR	Derived from the ATM Traffic Descriptor IB or from Network Management
CDVT	Set via Network management procedures and determined on the basis of local customer premises equipment
MBS	The maximum burst size (in cells) that will be sent on this connection. Set via Network management or from the SETUP signalling.
SCR	This parameter is intended to be an upper bound on the average cell rate on the connection. It can be derived from the information on the SETUP message or from Network management.

**Table 2:** VBR parameters

PCR and CDVT must be specified for the  $CLP=0+1$  stream. SCR and MBS may be specified for either the  $CLP=0+1$  stream or for the  $CLP=0$  stream.

## Conformance definitions

The ATM Forum traffic management specification [18] has three conformance definitions for VBR (with no distinction made between rt-VBR and nrt-VBR). The differences between these conformance definitions lie in whether or not tagging is supported, and whether SCR/MBS are specified for the entire stream or just for the CLP=0 stream.

All three conformance definitions require the equivalent of GCRA( $T_{0+1}$ , CDVT) where  $T_{0+1}=1/\text{PCR}$  as one of the conformance tests.

### VBR.1

In addition to the PCR GCRA, this conformance definition requires the equivalent of GCRA( $T_{s,0+1}$ ,  $BT_{0+1}+CDVT$ ) where  $T_{s,0+1}=1/\text{SCR}_{0+1}$ . Given that there is no separate conformance definition for the CLP=0 stream, no tagging is possible and the CLR guarantees can apply *only* to the entire CLP=0+1 stream. In order to be conformant, a cell must pass both the PCR and the SCR GCRA's.

### VBR.2

In addition to the PCR GCRA, this conformance definition requires the equivalent of GCRA( $T_{s,0}$ ,  $BT_0+CDVT$ ) where  $T_{s,0}=1/\text{SCR}_0$ . This is a conformance definition for the CLP=0 stream, but for VBR.2 no tagging is allowed. The CLR guarantees apply *only* to the CLP=0 stream. The guarantees for the CLP=1 stream and the combined stream is not defined. In order to be conformant, a cell with CLP=0 must pass both the PCR and the SCR GCRA's. A cell with CLP=1 need only pass the PCR GCRA in order to be conformant.

### VBR.3

This is the most flexible of the 3 VBR conformance definitions. In addition to the PCR GCRA, this conformance definition requires the equivalent of GCRA( $T_{s,0}$ ,  $BT_0+CDVT$ ) where  $T_{s,0}=1/\text{SCR}_0$ . This is a conformance definition for the CLP=0 stream. If tagging is supported by the network and requested by the user, then the network may tag cells that conform to the PCR GCRA and not to the SCR GCRA. The CLR guarantees apply *only* to the CLP=0 stream. The guarantees for the CLP=1 stream and the combined stream is not defined. In order to be conformant, a cell with CLP=0 must pass both the PCR and the SCR GCRA's. A cell with CLP=1 need only pass the PCR GCRA in order to be conformant.

## ABR

### Introduction

ABR uses closed loop control to implement flow control. Based on the information from the control loop ABR sources adapt their cell rate. Other service categories, i.e. CBR, VBR and UBR, use an open

loop. They use a traffic contract that is specified at connection setup time. They use the parameters of the traffic contract to determine their rate, and do not adjust their rate based on information received from the network.

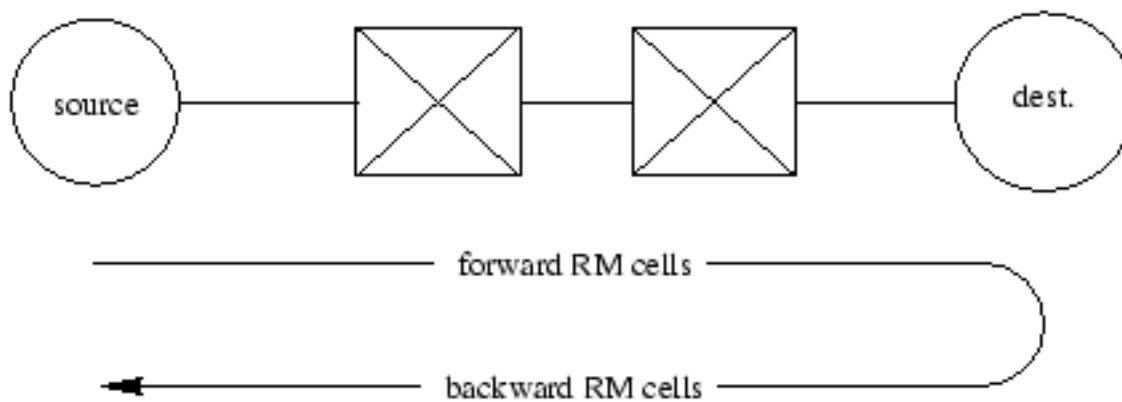
The point of ABR is that sources that do not need hard service guarantees can share excess bandwidth and adapt to bottlenecks in the network.

We will explore how control information is conveyed through the network and how control information is used to establish rate adaptation. It is primarily based on [18], [16] and [12]

## The basics

### The Control Loop

ABR uses feedback to establish a control loop from source to destination. Its basic operation is very simple: the network nodes are allowed to modify cell headers or specific cells called resource management (RM) cells. The source sends RM cells, which are possibly modified by the network. The destination turns around the modified RM cells and sends them back to the source. The network is once more allowed to modify the RM cells, thus closing the loop. The sources use the feedback information in the RM cells to determine at what rate they are allowed to send cells. Figure 1 shows a control loop that is terminated by two ATM end systems.



**Figure 1:** Unidirectional ABR control loop

### Resource Management Cells

RM cells are used to transport control information from source to destination and back. RM cells are distinguished from data cells by the SDU field in the header of an ATM cell. It is set to 6.

There are a number of different cells associated with an ABR connection:

- Data cells carry user data;



- Forward RM cells are sent by the source. They are subject to change by network nodes;
- Backward RM cells are 'turned around' by the destination and are sent back to the source to convey the feedback information to the source. Network nodes are allowed to change backward RM cells.

All cells associated with an ABR connection can be split into two categories: in-rate and out-of-rate cells.

- In-rate cells have to be sent within the current rate of the source. In other words they have to conform to the source's Actual Cell Rate (ACR). In-rate cells have CLP=0. In-rate cells can be user cells, forward and backward RM cells.
- Out-of-rate cells do not keep to the ACR. They may not be sent at a rate higher than the Tagged Cell Rate (TCR), which is set to 10 cells/s. In other words, out-of-rate cells do not have to wait for the next in-rate cell slot, but they may not be sent at a rate greater than TCR. Out-of-rate cells have CLP=1. Out-of-rate cells can only be RM cells.

In addition to in-rate/out-of-rate forward and backward RM cells there are also Backward Explicit Congestion Notification (BECN) cells. *BECN cells* are non-source generated cells; they are generated by a destination or switch and are sent towards the source. BECN cells allow the destination or the network to inform the source of congestion events that require immediate action. They need not wait for the next RM cell from the source to be turned around.

All RM cells have the same format, the different types are distinguished by different bits within the RM cell.

The fields that are important in the process of controlling the rate are shown in Table 3.

CI	Congestion Indication	1 bit
NI	No increase	1 bit
ER	Explicit rate	16 bit floating point <sup>1</sup>

**Table 3:** RM cell fields used for rate calculation

Note that the floating point number used for ER is a non-standard format, e.g., it is non-IEEE compliant.

### The Three Flavours of ABR

TM 4.0 describes three ways for a switch to indicate that a source is allowed to increase or must decrease their rate: Explicit Forward Congestion Indication (EFCI), Relative Rate (RR) and Explicit

Rate (ER) marking. In this section we will discuss them one by one. However, any of these methods may be implemented by a switch manufacturer and consequently all three may be present simultaneously in any given network.

### EFCI marking

When a node in the network that it is congested or that congestion is imminent, it indicates this by setting the EFCI bit in the cell header of cells in its queue. The destination stores the EFCI bit and reflects the EFCI state by copying it to the CI bit of a RM cell. When the RM cell with its CI bit set reaches the source it will have to decrease its rate with a factor RDF, the Rate Decrease Factor. The RDF defaults to  $2^{-16}$ . However, a source is always allowed to send at the minimum cell rate (MCR), so if the calculated rate is smaller than the MCR it is replaced by the MCR. (source rule 1, 8)

If a backward RM cell arrives without its CI bit set, the source is allowed to increase its rate by a fraction of the PCR. The size of the fraction is set by the Rate Increase Factor (RIF), which defaults to 1. The newly calculated ACR may not be higher than the PCR. (source rule 1, 8) Summarizing, we can calculate the new ACR from:

$$\text{ACR} \quad \leftarrow \begin{cases} \min(\text{ACR} + \text{RIF} * \text{PCR}, \text{PCR}) & \text{CI} = 0 \\ \max(\text{ACR} (1 - \text{RDF}), \text{MCR}) & \text{CI} = 1 \end{cases} \quad (1)$$

Note that by using the default RIF the ACR will always be set to the PCR, every time an increase is allowed the ACR will return to the PCR. This may cause oscillating behaviour on the part of the source, which may have a poor overall effect on the network.

### Relative Rate Marking

In comparison with EFCI marking, Relative Rate marking doesn't reflect the network state by altering the header of the cells, but uses the forward RM cells. In addition to setting the CI bit when the network is congested, a network node can also influence the sources behaviour by setting the No Increase (NI) bit of the RM cell. If a source receives a RM cell with  $\text{NI} = 1$  and  $\text{CI} = 0$ , it keeps sending at its current ACR. (source rule 8)

$$\text{ACR} \quad \leftarrow \begin{cases} \min\{\text{ACR} + \text{RIF} * \text{PCR}, \text{PCR}\} & \text{CI} = 0 \quad \text{NI} = 0 \\ \text{ACR} & \text{CI} = 0 \quad \text{NI} = 1 \\ \max\{\text{ACR} (1 - \text{RDF}), \text{MCR}\} & \text{CI} = 1 \quad \text{NI} = + \end{cases} \quad (2)$$

### Explicit Rate Marking

This third flavour of ABR uses a mechanism where the rate is explicitly specified by the network. The source sends out RM cells with the ER field set to the PCR of the connection. The network nodes are allowed to decrease the value of the ER field to reflect the rate they are able to carry. The idea is that the 'bottleneck' node of the connection will determine the new rate of the source. The destination turns the forward RM cells around. Once the source receives the RM cell, it copies the ER value to the ACR, with two restrictions: the ACR is not allowed to increase with a factor larger than RIF and b) the ACR must be larger or equal to the MCR. (source rule 1, 8 and 9)

Note that in networks where all nodes have implemented ER, CI and NI will always be 0 and since a forward RM cell starts with ER = PCR, a backward RM cell can never have a ER larger than PCR.

In short:

$$ACR \leftarrow \max\{MCR, \min\{ACR + RIF * PCR, ER\}\} \quad (3)$$

### Combining EFCI, RR and ER marking

When a network consists of a combination of all three types of ABR, the calculation of a new ACR is determined by the marking method of the 'bottleneck' node. This can be easily seen by supposing that one of the nodes along a connection is congested and replacing the congested node by all three flavours of ABR. In this case the formulae can be combined to:

$$ACR \leftarrow \begin{cases} \min\{ACR + RIF * PCR, ER\} & CI=0 \quad NI=0 \\ \max\{\min\{ACR, ER\}, MCR\} & CI=0 \quad NI=0 \\ \max\{\min\{ACR(1-RDF), ER\}, MCR\} & CI=1 \quad NI=+ \end{cases} \quad (4)$$

### Some details

In the previous section we have looked only at how a source is supposed to adjust its rate depending on the information that is sent back to it by means of backward RM cells. There still are a number of topics left to be discussed: signalling, the scheduling of data cells and RM cells, bidirectional ABR connections, ABR VPCs.

First we will look into the signalling of ABR connections and the initialization of them.

### Signalled parameters

We have already come across 4 parameters that are signalled: PCR, MCR, RDF and RIF. All are

negotiable. (Most of the ABR parameters are negotiable, but we will confine ourselves to an explanation of what the parameters do.) Except for the PCR, these parameters have a default value. For RDF and RIF we have already mentioned the default value. The MCR defaults to 0.

In addition to these three more parameters are signalled:

### **Initial Cell Rate (ICR)**

This defaults to the specified PCR and denotes the rate at which the source is allowed to send cells initially, or after a silent period.

### **Transient Buffer Expose (TBE)**

This parameter specifies the number of cells a source is allowed to send after a silent period, before the first RM cells arrives back at the source.

### **Fixed Round-Trip Time (FRTT)**

It is a measure for the round trip delay of RM cells. It consists of the sum of (fixed) switching, processing and propagation delays. From it another value of ICR can be calculated by seeing that it makes no sense sending TBE cells faster than FRTT, so:  $ICR = \min(ICR_{sig}, LBE/FRTT)$

The above three parameters have as a goal the limitation of initial surges by the sources. The idea is similar to TCP slow start in a way. These parameters limit the amount (and rate) of information that can flow into the network before the network has a chance to react with information in RM cells. In the TCP case, this limitation is extreme at first (i.e. 1 packet) and built up gradually. In the case of ABR, this limitation is expressed in the above parameters so that the network knows in advance how much data will be sent by a connection before feedback takes effect.

Optionally 4 additional parameters may be signalled: Nrm, Trm, CDF, ADTF. These four parameters attempt to ensure that the feedback loop remains in place and that the information on which both the network and the source depend is current. In particular, the cutoff decrease factor (CDF) ensures that in case too many RM cells are lost, the source reduces its rate so that the network remains protected.

### **Nrm**

is the maximum number of cells a source may send between each forward RM cell. Defaults to 32.

### **Trm**

is an upper bound on the time between two RM cells for an active source. When a source is sending at a low rate ACR a situation may occur where  $Trm < Nrm * ACR$ . Trm makes sure that the network is sensed at least every Trm seconds. It defaults to 100 ms.

### **Cutoff Decrease Factor (CDF)**

is a value similar to the RDF. A source can have CRM outstanding RM cells, where CRM denotes a maximum in the difference in sent and received RM cells. If the difference exceeds this value a source must decrease its rate. CRM is calculated as follows  $CRM = \lceil TBE/Nrm \rceil$ .

### **ACR Decrease Time Factor (ADTF)**

is the time permitted between sending RM cells before the source has to return to the ICR. In

essence, this is the time before a source is declared silent. This is to prevent the source from suddenly surging to the last ACR. ADTF defaults to 0.5 seconds.

### Initializing an ABR cell stream

There have to be a number of rules that apply to the situation where a ABR connection first sends cells. A more-or-less similar situation occurs after a silent period, i.e., the connection was up, but the source had no data to send, i.e. the ADTF time is up.

The ABR source sets its ACR to at most ICR, starts off the cell stream with an RM cell and sends at most TBE in-rate cells at the signalled or calculated ICR. (source rule 2) After the first RM cell is received it is `business as usual', since the feedback loop at this point is fully operational.

A source sends all RM cells with information fields set to the initial value as given by Table 5-4 in the [18], i.e. EFCI, CI, NI are reset (source rule 10 and 12) and ER is set to the PCR, or optionally an estimate of the rate it will need. The latter case is an example of a use-it-or-lose-it policy. (source rule 13) We will get to that later.

### Modes of Operation

We distinguish a 3 modes of operation to explain the use of forward RM cells by the source: Normal, slow and silent.

#### Normal mode of operation

We define the normal mode of operation as one where the source always has data cells to send. The network is the bottleneck in this case. At first we assume that the ACR is high enough to not time out various timers, such as ADTF and Trm.

The source sends the data cells interleaved by forward RM cells. After sending at most  $N_{rm}-1$  data cells it has to insert a RM cells into the stream. This means that a RM cell is send every  $N_{rm}/ACR$  seconds, provided the ACR did not change during this interval. We have encountered this boundary before when we explained the Trm parameter.

When a source receives a RM cell, it uses Eq. 4 to calculate its new ACR. From that moment on data and in-rate RM cells are obligated to follow the ACR.

#### Slow mode of operation

Suppose that a source has only a few to send, or the ACR is set low because of heavy network congestion. In this case the intervals between the forward RM cells get larger. A result of this is that the source doesn't receive network feedback very often. To keep in touch with the network, a source starts

sending RM cells if the time since the last RM cell is larger than  $Trm$ , which defaults to 100 ms. If the ACR is high enough it sends a in-rate RM cell, else it may send an out-of-rate RM cell. Note that the minimum interval for out-of-rate RM cells is 100 ms, the same as  $Trm$ .

### Silent mode of operation

It can happen that even though a source has  $ACR > 0$ , it does not have any cells to send. In this case, according to source rule 3 in [18], the source may not be able to send any in-rate RM cells. This occurs because although there is a maximum time between RM cells ( $Trm$ ) there is also a parameter governing the minimum number of data cells that must be sent between RM cells ( $Mrm$ ). If this parameter is non-zero and the source is quiescent (no cells to send) then eventually the ADTF timer will expire and the source will reduce its rate to  $\min\{ICR, ACR\}$ .

### Missing RM cells

Due to congestion within the network, cells can be lost. If RM cells are lost, this means that the source does not get network feedback. In order to force the source to reduce its rate in such a situation, a source calculates the difference in sent and received RM cells. If the difference is larger than  $Crm$ , the source has to reduce its rate by a factor CDF. The new rate is calculated from:

$$ACR \leftarrow \max\{ACR * (1 - CDF), MCR\} \quad (5)$$

### Cell Scheduling of In-Rate Cells

At any time a source can have three kinds of cells to send: data cells, which are always in-rate, forward and backward RM-cells. According to the rules of source behaviour, forward in-rate RM cells have priority if the minimum timing conditions for forward RM cells are met. If a forward in-rate RM cell is not due to be sent, and a backward RM cell is ready, then the backward RM cell can be sent in-rate if it is the first one since the last forward in-rate RM cell or if there are no data cells waiting for transmission. If neither a forward nor a backward RM cell is to be transmitted, then a data cell can be transmitted if one is ready.

Note that backward RM-cells are counted as in-rate cells and thus may trigger more forward RM-cells than expected from the number of data cells sent by the source. We have already come across the implications of source rule 3a, but now we see that it also has implications for the behaviour of the destination.

In the uni-directional case, the destination does not send data cells, but does turn around forward RM-cells. These backward RM-cells can be scheduled for in-rate transmission, thus source rule 3a applies to them.

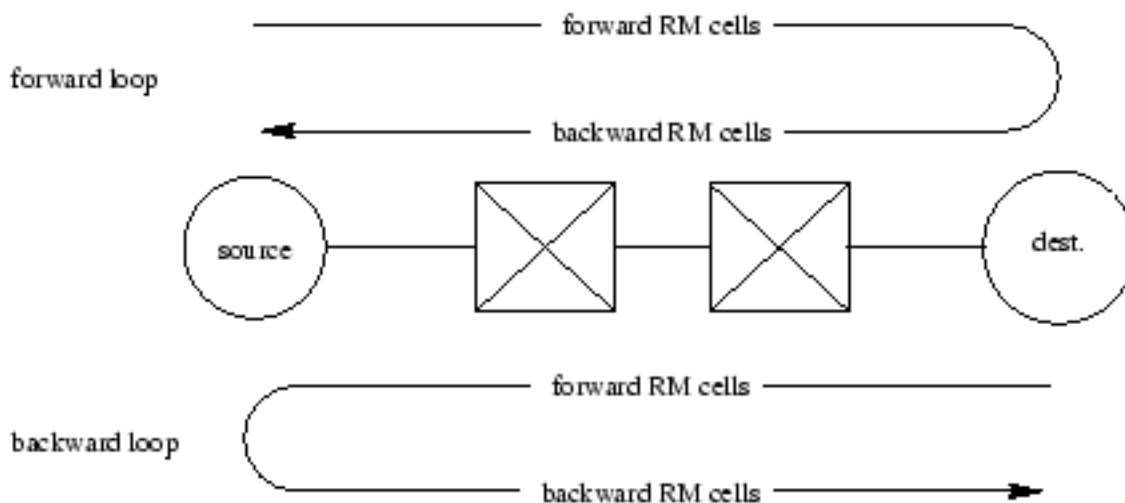
The cell rate for in-rate backward RM-cells is at most  $1/32$ nd of the ACR of the source, or else  $1/T_{rm}$  cells/s for slow source. A silent source does not send any in-rate RM cells, so the in-rate backward RM-cell rate is 0.

If the network is congested in the direction from destination to source, it is possible that backward RM-cells are lost, due to, for instance, buffer overflow. In this case the outstanding RM-cell count will exceed  $C_{rm}$  and the source must reduce its ACR according to Eq. 5.

There is one question still open: does a destination also use a forward RM-cells to sense the network, even if it has no data cells to send. In other words, does it open a control loop to calculate the rate at which it is allowed to send turned around RM-cells.

Appendix I.7 makes an analysis of a number of situations, among which is a case where the backward  $ACR = 0$ . It states that all backward RM cells are sent out-of-rate or not at all. This implies that the source is not informed as often as it should and now has two options: Reduce the number of RM cells so that CRM is satisfied, or risk cut-off because of missing RM cells.

For bidirectional data transfers, this isn't an issue. Both sides should monitor the network and continually recalculate the ACR. Because of this both sides should issue RM cells. Now there's certainly a double control loop.



**Figure 2:** Bi-directional control loop

### Virtual Source/Virtual Destination

A control loop may be segmented into two or more. This is done by VS/VD Control. The idea is that a network node splits itself into two parts: a virtual destination on the side of the source and a virtual source on the side of the destination. This process may repeat a number of times along a connection.

## ABR VPCs

The mechanisms for ABR VPCs are the same as for ABR VCCs, with the expectation that all RM cells belonging to the VPC are sent using  $VCI = 6$ .

### Use-it-or-Lose-it policy

A use-it-or-lose-it policy means that if the source transmits at a rate significantly below the ACR, then the ACR is reduced to the (approximate) real source rate. The source itself or the switches may implement this policy.

The goal of this kind of behaviour is to keep the ACR close to the actual transmission rate of a given source. From the network point of view, use-it-or-lose-it behaviour is advantageous because it prevents sudden traffic surges and it allows switches to use bandwidth that would otherwise be unused. However, there are disadvantages from the source point of view. Specifically, if there is a traffic surge from the source, it may have to buffer the traffic for one or more round-trip times until the ACR is increased again. Given current conditions at the bottleneck link, it may not be possible to increase the ACR at the time that a connection wants it. Further, sources that do *not* implement use-it-or-lose-it behaviour have an advantage over sources that do implement it. This behaviour would be fairer if implemented by the network switches.

## UBR

Unspecified Bit Rate (UBR) is the lowest level of ATM Service category. There is no QoS guarantee for connections using this service category and any flow control is assumed to take place on a higher level (i. e. the 'application' level). UBR is intended for use by traditional data communications applications where the end-to-end delay can be highly variable and the packet loss can (occasionally) be high.

UBR, though it receives no guarantees, still requires a source and connection traffic descriptor. Specifically, the UBR source *must* specify PCR as the source traffic descriptor, and the CDVT must also be specified (this can be done through network management procedures, for example, or can be assumed by the network). The PCR applies to  $CLP=0+1$  traffic. The ATM Forum, however, says that the network is not obliged to use the connection traffic descriptor either for connection admission control or for usage parameter control. However, the network may reject a UBR connection on the basis of the PCR source traffic descriptor, or the PCR requested may be negotiated to a lower value by the network.

There are two conformance definitions for UBR, one which does not include tagging and one which does. The first conformance definition, named UBR.1 says simply that the network may not change the Cell Loss Priority bit in any cell of the connection. A GCRA based on PCR and CDVT is not specified. The second conformance definition, UBR.2 simply specifies that *any* cell may have its CLP bit changed to 1. There are no rules given for doing so, however. With the tagging option, the tagging of a cell does not imply that the cell is non-conforming. One must conclude that any *arbitrary* cell can be tagged on a UBR connection with the UBR.2 conformance definition.



Since there is no QoS guarantee for UBR connections, they can be treated (and multiplexed) in any convenient way by the network. In particular, it is not required that UBR connections sharing a link get a fair share of the bandwidth of that link. They can be scheduled in any way that is convenient to the implementation. Some implementations may choose to try to give UBR connections fair access to bandwidth, but that is not strictly required.

## Discussion

UBR has the simplest specification of all the ATM service categories except for CBR. However the cost is complete lack of guarantees. As shown in [17] unless some more sophisticated way of discarding cells from packets is implemented, serious throughput degradations can occur when UBR is used. Even if EPD and/or PPD are implemented in switches, unfairness can occur among UBR connections sharing the same link.

UBR is nonetheless still appropriate for traditional data traffic when used properly and when EPD/PPD are implemented in the intervening switches.

## GFR

### Introduction

Guaranteed Frame Rate (GFR) is the newest of the service classes defined by the ATM Forum. It was originally introduced by Heinanen and Guerin [8] as a modification of UBR called UBR<sup>+</sup>. The purpose of defining a new service category was simple and clear: Users, particularly data users, have no idea in general of the detailed characteristics of their traffic. However, they still view it as desirable to have some form of QoS guarantee, as opposed to using UBR service which gives no guarantees whatsoever. GFR is an admission, to some extent, that ABR is a complex design to implement and that it will not be fully deployed in the near term by all equipment.

The basic technical idea of GFR is that even in the absence of a complete traffic characterization, something can still be known about a source. In fact enough can be known to characterize the source in a very *loose* sense, so that internally in the network, resources can be reserved for the source to guarantee better service than UBR.

### Original Proposal

The original proposal for GFR was made in the ATM Forum contributions [8, 9]. The proposal consists of several parts: informal and formal service definitions, with a discussion of conformance and tagging; pseudo-code showing sample implementations; and a set of criteria for evaluating the service.

### Service definitions

The GFR user must specify a maximum packet size that he will submit to the ATM network and a minimum throughput that the user would like to have guaranteed. The user may send packets in excess of this guaranteed service rate, but they will be delivered on a best effort service. If the user remains within the throughput and packet size limitations, he can expect that the rate of packet loss will be very low. If the user sends in excess of the guaranteed rate, he can expect that if resources are available, they will be shared equally among all competing GFR users.

Under the original proposal, GFR service is specified by a combination of explicit and implicit parameters. The cell level traffic parameters are given in Table 4.

Traffic Parameter	How derived
PCR (CLP=0+1)	Derived from the ATM Traffic Descriptor IB or from Network Management
CDVT	Set via Network management procedures and determined on the basis of local customer premises equipment
MCR	Source not clearly identified. Assumed to be specified via Network management.
MBS	Negotiated via Network management or via the AAL CPCS-SDU size parameter of the SETUP message. Set to 2*CPCS-SDU in units of cells
BT	Derived from MBS

**Table 4:** GFR parameters

Given these parameters, the following rules are used for GFR:

### Conformance

This is determined by the following rules:

1. The PCR GCRA. Non-conforming cells may be tagged or discarded (but see rules for this below).
2. The maximum packet size.

### Guarantees

These apply to user traffic that fulfills the following conditions:

1. Passes a modified GCRA with the parameters: GCRA(1/MCR, BT(MBS)+CDVT).
2. GCRA modifications:

When the first cell of a packet arrives

If GCRA token count  $\geq$  MBS/2

Then *All* cells of the packet are eligible for service

and each cell consumes a token when forwarded.

Else No cell of the packet is eligible for service

and none consumes a token when forwarded.

Fi

Note that an implementation may tag cells which do not meet the eligibility criteria.

## Expectations

There is the expectation that the excess traffic (i.e. the traffic not receiving guarantees, but still conforming) will be delivered on the basis of a *fair share* of the available resources of the network at the time.

## Tagging

There are two rules regarding tagged cells:

1. Tagged cells (having been tagged for whatever reason) are not eligible for service guarantees and therefore do not 'pass through' or trigger mechanisms (such as the modified GCRA) designed to determine eligibility.
2. All cells from the same packet must be tagged identically.

There are several important points to note in the description of GFR service: First, the GFR node must be aware of the packet level boundaries of the user input stream. For AAL-5 packets this is a fairly simple matter (looking at the control bits in the ATM header), but for other AALs, (and in particular for future AALs) this may not be so simple. Second, the service guarantees are held to apply to a number of cells that are deemed eligible for the guarantee, but not necessarily to the exact cells that pass the modified GCRA. This is a matter of flexibility, but also incurs some risk. In this interpretation, if the network delivers some ineligible cells and at some later time discards eligible cells, then the guarantee is still met. Third, the requirement to offer a 'fair share' of resources for the delivery of cells sent in excess of the guarantee *requires* that there be some mechanism in the network nodes which provides this controlled sharing. The authors suggest Weighted Fair Queueing (WFQ)[[1](#)] as an example of such a mechanism. One of the points of contention during the development of this item is whether it is even possible for a service discipline such as FCFS to deliver a 'fair share' of resources.

## Current Proposal

The GFR definition is still in flux. There are a number of proposals still outstanding and at the December 1997 meeting in Singapore, they reached a compromise (which is not yet published). In this section, we outline the latest GFR definition by one of the original proposers and some others [[10](#)], and discuss some of the problems pointed out by others in other contributions [[20](#), [22](#), [21](#)].

## Informal Service Definition

The informal service definition is much the same as the original with a few points sharpened. The user must specify the parameters shown in Table 5 (through signalling or subscription).

Traffic Parameter	How derived
PCR (CLP=0+1)	Derived from the ATM Traffic Descriptor IB or from Network Management
CDVT	Set via Network management procedures and determined on the basis of local customer premises equipment
MCR	Source nor clearly identified. Assumed to be negotiated via Network management.
MBS	Assumed to be negotiated via Network management.
MFS	Again, Assumed to be negotiated via Network management, and specifying the largest frame (in numer of cells) that will be sent on the connection.
BT	Derived from MBS, PCR and MCR

**Table 5:** GFR parameters (2nd attempt)

The service guarantee applies to traffic streams which stay within the given parameters and says that the user can expect the frames to be delivered "with minimum losses". The traffic which is sent in excess of the Minimum Cell Rate (MCR) and Maximum Burst Size (MBS) is delivered within the limits of available resources and each connection should get a "fair share" of those resources. The user can send both marked (i.e. CLP=1) and unmarked (CLP=0) traffic, but the service guarantee applied *only* to CLP=0 traffic. The user can request tagging from the network, and in this case, the network may tag cells of CLP=0 frames either if they do not conform, or if they are not eligible for the service guarantees.

## Formal Service Definition

As with the earlier definition, the authors distinguish between conformance rules and service guarantee rules. There are different processes that identify whether a frame conforms and whether a frame can get service guarantees. As with other service definitions, the service guarantees do not apply to cells of specific frames, but only to the *number* of cells of frames that were determined eligible. Thus, cells from frames which conform and are eligible may be discarded and those from frames which have been tagged may be delivered. It is important to note that in [10] the service guarantee is given in terms of *cells*.

## Conformance

Conformance is defined in terms of cells (not frames). A cell conforms if

1. The cell passes GCRA(1/PCR, CDVT).
2. The cell does not cause the frame to exceed the MFS as defined above.

If a cell does not conform, according to the above rules, the reaction of the network depends on the position of the cell. If it is the first cell of a frame, the network can discard or tag the whole frame. If it is not the first cell of a frame, then the network can discard the rest of the frame, but may not tag the rest of the frame. It can also, of course simply send the remaining cells onward. In case the network decides to discard, it must not discard the last cell of the frame since that contains the frame boundary. Service Guarantee Eligibility

The mechanism for deciding whether the cells from a frame are eligible for guaranteed service is a modified GCRA (the F-GCRA). The modified GCRA has two parameters  $1/MCR$  and  $\tau = BT + \tau_{MCR}$ . The burst tolerance,  $BT$  is defined as  $(MBS-1) \cdot (1/MCR - 1/PCR)$ . The modified GCRA is somewhat different from that first proposed. Unmarked (and untagged) cells that pass this GCRA test may or may not be delivered, but at least an equal number of cells is expected to be delivered. The network is permitted to perform whole frame tagging (or discarding), or it can perform partial frame discarding of frames which are not eligible for service guarantees.

## F-GCRA

The F-GCRA for testing eligibility is defined as follows:

Initial values:  $X = 0$ ,  $LCT = t_0$

When the first cell of a frame arrives

$X_t = X - (t_0 - LCT)$

If  $(X_t > \tau)$  OR (cell is tagged)

Then Cell is not eligible

Else Cell and Frame are eligible

Fi

If (eligible)

Then  $X_t = X - (t_0 - LCT)$

$$X = \max\{0, X_t\} + T$$

$$LCT = t_0$$

Fi

For remaining cells of a frame which are not discarded due to non-conformance:

If (eligible)

$$\text{Then } X_t = X - (t_0 - LCT)$$

$$X = \max\{0, X_t\} + T$$

$$LCT = t_0$$

Fi

In the above,  $T$  is the first parameter of the leaky bucket (i.e.  $1/MCR$ ) and  $\tau$  is the second (i.e.  $BT + \tau_{MCR}$ )

[Back to list of results](#)

## 2. Test concatenation of networks with different ATM service categories

In case the networks do not have the same service categories (like in WAN->CBR/VBR and in LAN->UBR/ABR), one has to investigate what will happen and how problems can be solved.

Such an experiment with proprietary ABR has been done between NL and CH to see what is possible with a kind of ABR (remember it was no ABR as defined by ATM Forum, so it is only done to get a feeling of the behavior!).

### Network configuration

UU-Physics borrowed a Gigaswitch from Digital Equipment for the concatenation tests. Three workstations were available for the tests. Two DEC Alpha workstations, were connected to the Gigaswitch ``test 1" in Utrecht. One workstation in Utrecht was configured to be *ARP* server for the entire cluster. The other Gigaswitch, ``test 2", was located in Geneva together with an DEC Alpha workstation. The Gigaswitch was connected directly to the Fore ASX 200 switch in at University of

Geneva. An overview of the network configuration of this test setup is given in figure [8.5.2.1](#).

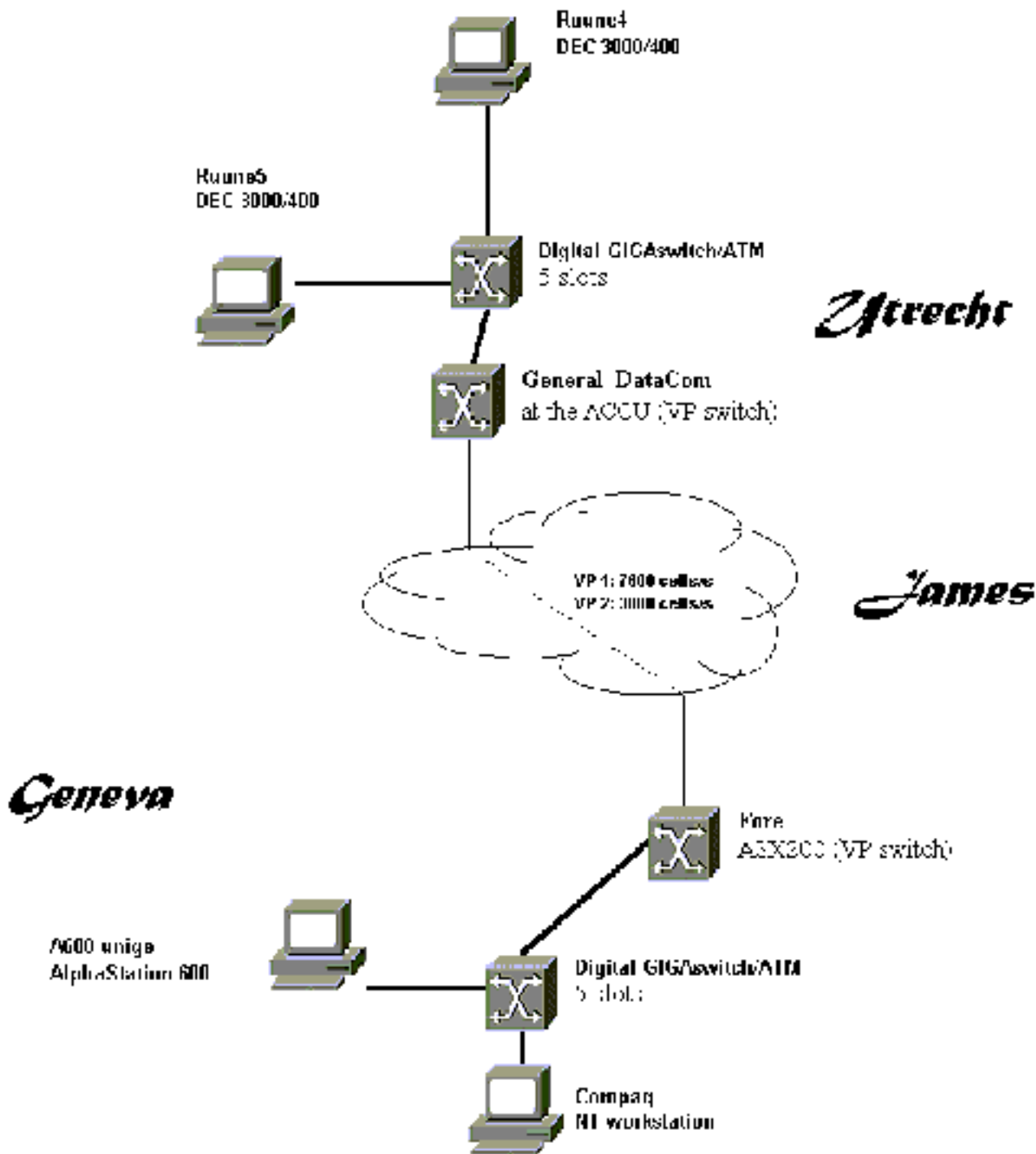


Figure 8.5.2.1 Network configuration of the international concatenation test.

### Measured throughput

Over the 7600 cells/s VPC, One Digital-ABR PVC was opened between one workstation in Utrecht and

Geneva. The other workstation in Utrecht used CBR SVC's to communicate with the Geneva workstation:

- Throughput was tested with programs like *bricks* and *UMAC*.

It appeared to be very close to the bandwidth setting of the tunnel, i.e. the measured bandwidth was at least 99% of the calculated 2.75 Mbit/s. Also the 3800 cells/s VP was tested, and found to be filled up to at least 99% of the calculated 1.38 Mbit/s.

- When two DEC workstations, that were connected to the Gigaswitch in Utrecht, communicated with the DEC workstation in Geneva, the bandwidth was divided as expected, i.e. 2 Mbit/s for the Digital-ABR PVC and the rest for the CBR SVC's.
- When the traffic on the CBR connection was stopped, the Digital-ABR connection took over the entire bandwidth.
- During all throughput tests Digital-ABR performed as expected, i.e. almost the full bandwidth was used and no cells were lost on either the Digital-ABR or the CBR VC.

## Combining VPC's

Another interesting question is how the switch at the endpoint of a tunnel handles traffic from other switches in combination with traffic coming from a workstation. For traffic coming from another switch, no flow control is available. Therefore it is interesting to know how the bandwidth is divided between the different sources.

We got two VPC's with a bandwidth of 7600 cell/s each, one from Utrecht to Utrecht and one from Utrecht to Geneva, see figure [8.5.2.2](#). Two each switch, one Digital workstation (with flow control) was connected. All traffic from Utrecht to Geneva was sent through Gigaswitch ``test 1''.

The flow control in the Digital Gigaswitch is implemented in such a way that each workstation is allowed to send at least 20 cell/s. Furthermore, if incoming traffic from two or more ports has to be combined on a (different) outgoing port, the buffers are read with a round robin mechanism.



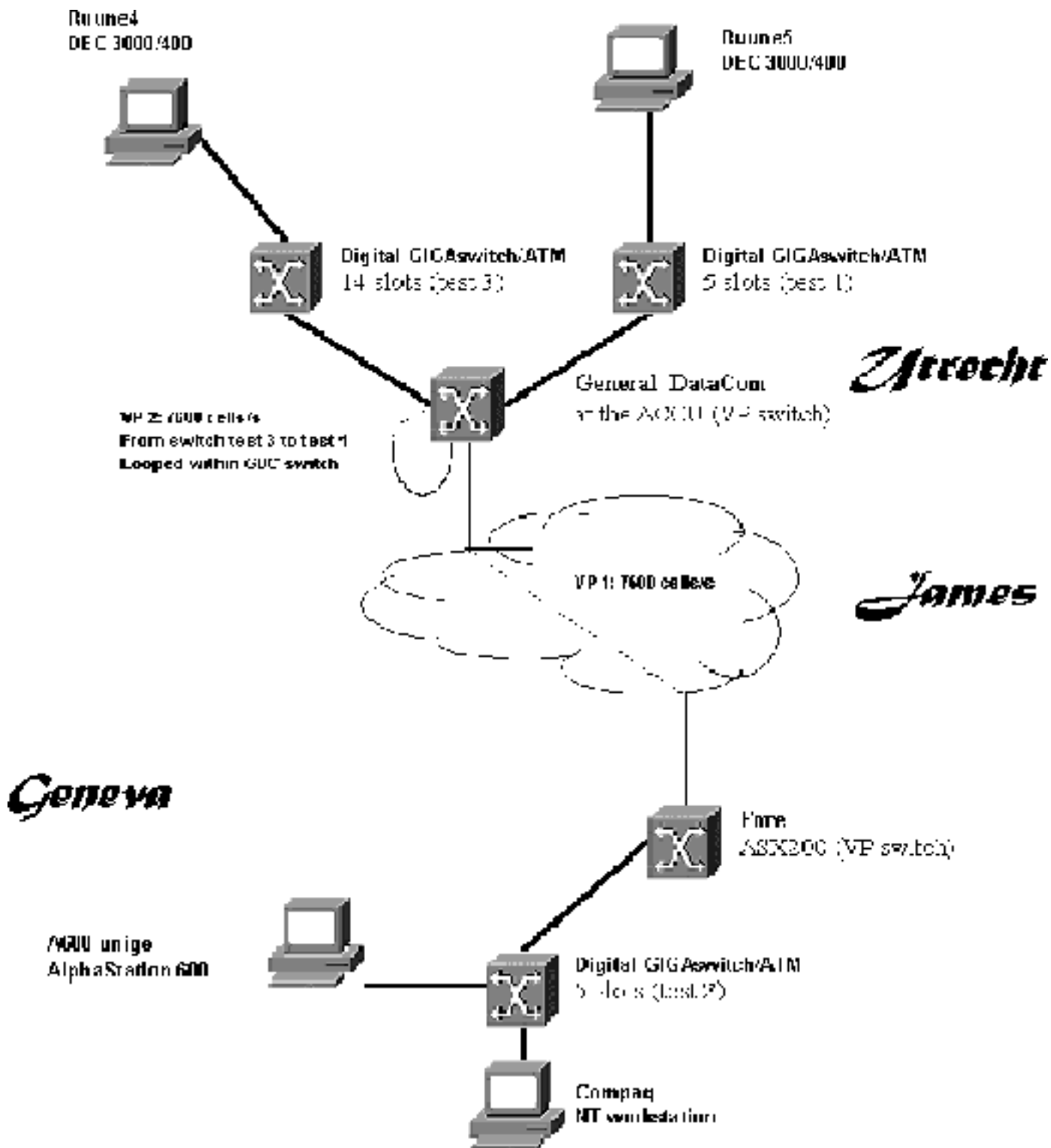


Figure 5.8.2.2 Network configuration of the international combination VPC test.

**TCP traffic**

First we tried sending TCP traffic from workstation ``ruune4'' to ``ruunf7''. The complete bandwidth was used and no cells were lost. Then we added traffic from ``ruune5'' to ``ruunf7''. We expected ``ruune5'' hardly to be able to capture any bandwidth as the traffic coming from Gigaswitch ``test 3'' could not be reduced. What happened was completely different. As the buffers are read with a round robin

mechanism, the Gigaswitch ``test 1" divided the bandwidth between the two sources by throwing away 50% of the cells coming from the other Gigaswitch (``test 3"). Workstation ``ruune5" was allowed to fill the other 50% of the tunnel. Initially this caused all traffic from ``ruune4" to ``ruunf7" to be lost. After a few seconds, the TCP layer reduced traffic and only a few cells coming from ``ruune4" were lost. This loss of cells was so low that after a few seconds constantly more than 49.9% of the bandwidth could be used by ``ruune4". A few seconds after the traffic from ``ruune5" was stopped, the traffic from ``ruune4" increased until the entire bandwidth of 7512 cell/s. was used.

All cell loss occurred in Gigaswitch ``test 1", as it's buffer with traffic coming from Gigaswitch ``test 3" overflowed. No cells were lost in the JAMES cloud, as both tunnels were configured with a bandwidth of 7600 cell/s.

### UDP traffic

Thereafter we did the same tests with UDP traffic. Again, the entire bandwidth was used, and no packets was lost when only traffic from ``ruune4" to ``ruunf7" was sent. Then, we started UDP traffic from ``ruune5" to ``ruunf7". Again, flow control secured that 50% of the bandwidth could be used by ``ruune5", and 50% of the packets coming from Gigaswitch ``test 3" were thrown away. As the UDP traffic does not decrease automatically, no traffic got through from ``ruune4" to ``ruunf7" until ``ruune5" stopped transmitting. Again, all cell loss occurred in Gigaswitch ``test 1" and no cells were lost in the JAMES cloud.

### Conclusion

The VPC tunneling support, as implemented in the Pre-Field-test version of the firmware v2.5 software for the Digital Gigaswitch works well in a concatenation environment. At least when at both ends of the tunnel a Gigaswitch is used. It was possible to setup SVC's over the VPC, from Utrecht to Geneva and vice versa. The proprietary flow control, as implemented by Digital Equipment, provides for the possibility to automatically divide the bandwidth, that is assigned to a VPC, over the various SVC's. This feature is very important as it provides for an automatic and user independent way to make optimal use of the available bandwidth. This can be done without slowing down local connections and without the occurrence of cell loss.

If traffic from more sources is mixed in one VPC. The bandwidth appears to be equally divided. This means that packets can be thrown away for sources that do not implement flow control. In case the traffic coming from a source without Digital-ABR is TCP traffic, the TCP layer reduces the bandwidth used until only a few ATM cells are lost.

[Back to list of results](#)

### 3. Test ABR in LAN and WAN

## Sub activities

In this project, the following sub activities were defined:

1. Gather information concerning TM 4.0 (ABR) support from NIC vendors
2. Gather this information from ATM-switch vendors
3. Test basic RM capability of (VP) switches

The first two activities were necessary to see which equipment could be obtained and used in the actual limited ABR activity.

## Network Interface Card vendors supporting ABR

After reviewing end 1997 specifications and consulting people from Digital Equipment, ATMLink, IBM, FORE, EntraNIC, Adaptec, Interphase, RapidFire, Soliton, FeatureNet, Efficient Networks and SMC, it appeared that Digital Equipment was the only vendor claiming hardware/firmware support for ABR (all three forms; EFCI/RR/ER), FeatureNet claimed RM cell processing in hardware (no explicit ABR support), FORE was issuing beta NICs capable of ABR ER. Digital Equipment replied to have yet (end 1997) no drivers for their own card supporting ABR (neither for NT nor UNIX).

## Switch vendors supporting ABR

5 different vendors were reviewed in end 1997:

- Cisco (LightsStream 1010)

Support for ABR RR/EFCI

- Digital Equipment (GIGAswitch/ATM)

No ABR support yet

- Fore Systems (ASX 200)

Support for ABR EFCI/ER

- General DataCom (APEX line)

No ABR support yet, claimed support in new hardware for ABR EFCI

- Newbridge (Vivid CS 3000)

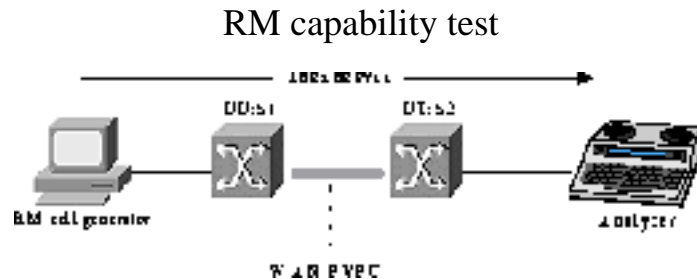
Support for ABR ER/RR/EFCI

## Basic RM capability tests

The equipment used in this basic RM capability experiment is:

- Two ATMWorks351 NICs from Digital Equipment
- ENI155p NIC from Efficient Networks
- Two Digital UNIX workstations
- Intel-based workstation, running both Windows NT and Linux
- Digital GIGAswitch/ATM ATM switch
- Cisco Lightstream 1010 ATM switch

The ATMWorks351 NICs from Digital Equipment have hardware support for ABR ER/RR/EFICI. No driver support for ABR flow-control was yet provided by Digital Equipment however, so these cards could not be used. The ENI155p NIC from Efficient Networks is not ABR capable at all, but is able to send raw (RM) cells when used in combination with freeware driver software for Linux. In this way, it is possible to send RM cells and see how switches deal with them by viewing egress RM cells with an ATM analyzer. In order to check if included switches in the next scenarios were able to (correctly) pass RM cells, either in- or out-of-band, the Linux workstation was used to assemble and send RM cells with different parameters. These RM cells were then sent over CBR PVPCs to an ATM analyzer over a cloud of ATM switches. The first test included University of Twente. To do so, a PVPC over the PTT Telecom ATM test network was acquired over which two switches at both Universities map a PVCC. The Linux end station will send it's RM cells over this PVCC to University of Twente:



The second test also included University of Twente, but the PVPC was first switched to Koln, Germany to the TF-TEN partner at University of Stuttgart (RUS). RUS switched one PVCC back to another VCI in the same PVPC which was then sent to University of Twente. The reason for this more elaborate test was to include as many switches as possible. All tests used CBR PVPCs of 1878 cells/s.

### RM cell contents

The below table contains the most important RM cell parameters.

**Table 5.8.3.1:RM cell parameters**

Field	Value	Hex value
Minimum cell rate	0 cells/s	0x00 0x00
Current cell rate	1878 cells/s	0x55 0xAB
Explicit cell rate	3756 cells/s	0x57 0xAB

The given cell rates are two bytes each, using the [encoding in table](#). The formula to calculate the cell rate from this table is:

$$R = 2^e \times (1 + m/512) \times nz$$

For example, for a cell rate of 1878 cells/s, the exponent would be 10 and the mantissa 427. This would result in two bytes: 01010101 and 10101011, 0x55 and 0xAB.

**Table 5.8.3.2:** Cell rate encoding

Bit(s)	Meaning
16	Reserved
15	Non zero value (nz)
14-10	Exponent (e)
9-1	Mantissa (m)

The CRC-10 code in each RM cell was generated using public domain software. All other fields contained default values. See [24] for details.

## Results

Both the first and second test were successful in the sense that RM cells were received in Enschede with the same information content as sent and correctly identified as valid RM cells. This means that no switch in the used paths blocked RM cells or modified its contents. The path including Koln, Germany consisted of these switches:

1. Digital GIGAswitch/ATM (Utrecht, NL)
2. General DataCom APEX (Utrecht, NL)

3. Lucent Globeview 2000 (Amsterdam, NL)
4. Siemens EWSXpress (Koln, DE)
5. Cisco Lightstream 1010 (Stuttgart, DE)
6. Siemens EWSXpress (Koln, DE)
7. Lucent Globeview 2000 (Amsterdam, NL)
8. General DataCom APEX (Enschede, NL)
9. Cisco Lightstream 1010 (Enschede, NL)
10. UB networks GeoSwitch155 (Enschede, NL)

[Back to list of results](#)

## References

- 1 Alan Demers, Srinivasan Keshav, and Scott Shenker. Analysis and simulation of a fair queueing algorithm. *Internetworking: Research and Experience*, 1(1):3-26, 1990.
- 2 B.T. Doshi. Deterministic rule-based traffic descriptors for B-ISDN: Worst case behaviour and connection acceptance control. In J. Labetoulle and J.W. Roberts, editors, *ITC 14*, pages 591-600. Elsevier Science bv, 1994.
- 3 Sally Floyd. TCP and explicit congestion notification. From LBL WWW page, 1996.
- 4 Sally Floyd and Van Jacobson. Random early detection gateways for congestion avoidance. *IEEE/ACM Transactions on Networking*, 1(4):397-413, August 1993.
- 5 Rohit Goyal, Raj Jain, Shiv Kalyanaraman, Sonia Fahmy, and Seong-Cheol Kim. UBR+: Improving the performance of TCP over ATM-UBR service. In *Proceedings of ICC '97*, Montreal, Canada, June 1997.
- 6 Rohit Goyal, Raj Jain, Shiv Kalyanaraman, Sonia Fahmy, Bobby Vandalore, and Sastri Kota. TCP selective acknowledgements and UBR drop policies to improve ATM-UBR performance over terrestrial and satellite networks. In *Proceedings of ICCCN '97*, Las Vegas, Nevada, USA, September 22-25 1997.
- 7

Rohit Goyal, Raj Jain, Shiv Kalyanaraman, Sonia Fahmy, Bobby Vandalore, Sastri Kota, and Pradeep Samudra. Simulation experiments with guaranteed frame rate for tcp/ip traffic. ATM Forum Contribution: 97-0607, September 1997.

8

Roch Guérin and Juha Heinanen. UBR+ service category. ATM Forum Contribution 96-1598, December 1996.

9

Roch Guérin and Juha Heinanen. UBR+ enhancements. ATM Forum Contribution 97-0015, February 1997.

10

S. Jagannath, N. Yin, J.B. Kenney, J. Heinanen, J. Axell, and K.K. Ramakrishnan. Modified text for guaranteed frame rate service definition. ATM Forum Contribution 97-0883, December 1997.

11

John B. Kenney. Satisfying UBR+ requirements via a new VBR conformance definition. ATM Forum Contribution 97-0185, February 1997.

12

Kerry W. Fendick. Evolution of Controls for the Available Bit Rate Service. *IEEE Communications Magazine*, pages 35-39, November 1996.

13

Siavash Khorsandi. UBR+ as enhanced UBR: Proposed amendments to the TM living list. ATM Forum Contribution 97-0361, April 1997.

14

M. Laubach. RFC1577: Classical IP and ARP over ATM. Technical report, IETF, January 1994.

15

Surya K. Pappu and Basak Debashis. TCP over GFR implementation with different service disciplines: A simulation. ATM Forum Contribution 97-0310, April 1997.

16

Raj Rain, et. al. Source Behavior for ATM ABR Traffic Management: An explanation. *IEEE Communications Magazine*, pages 50-55, November 1996.

17

Allyn Romanow and Sally Floyd. Dynamics of TCP traffic over ATM networks. *IEEE Journal on Selected Areas in Communication*, May 1995.

18

Shirish Sathaye (Ed.). Traffic management specification 4.0. ATM Forum, April 1996. af-tm-0056.000.

19

S Shenker, L. Zhang, and D. Clark. Some observations on the dynamics of a congestion control algorithm. *ACM Computer Communications Review*, 20(4):30-39, 1990.

20

Robert Wentworth. Issues related to the GFR service definition. ATM Forum Contribution 97-0922, December 1997.

21

Robert Wentworth. Supplemental text for the gfr service definition. ATM Forum Contribution

97-0980, December 1997.

**22**

Robert Wentworth, Deepak Kataria, and Tom Worster. Updated text for the GFR service definition. ATM Forum Contribution 97-0954, December 1997.

**23**

L. Zhang and D. Clark. Oscillating behavior of network traffic: A case study simulation. *Internetworking: Research and Experience*, 1:101-112, 1990.

**24**

The ATM Forum. *1996 Traffic Management Specification af-tm-0056.000*.



## 5.9. ATM Address Resolution

### Experiment leaders

Vegard Engen, UNINETT(BDC) and Olav Kvittem, UNINETT, Norway

### Summary of results

We managed to build a core network of 7 routers connected to the same ATM-cloud, spanning 6 countries. We tested both static routing and OSPF on this network. Tests showed that NHRP can work well on a core backbone, and within a network with loaded routers and switches it might be useful to bypass routers.

The ability to decrease the actual roundtrip time for each packet was limited by the fact that we had no native signaling, and had to tunnel the signalling info on our overlay network. Nevertheless, just decreasing the number of entries and exits on routers and switches becomes useful when the network reaches a certain load.

However, the tests did show that the mechanisms worked well between routers. There were some problems with making the traffic from the workstations bring up the shortcuts in a consistent manner. Sometimes it would take a long time before the shortcut was set up, and at the time of writing the reason is still unknown.

NHRP worked very well in an OSPF-environment. For this to work in a system where the total bandwidth is limited, having ABR would be useful to prevent smaller traffic loads established on an earlier time from getting a too large network capacity reserved compared to larger traffic loads reserved later. RSVP would also be of some value.

The original plan was to perform MPOA experiments too. However, we were not able to do those experiments due to lack of implementations of that protocol.

### Participants

- Vegard Engen (BDC), UNINETT (Norway)
- Robert Stoy, DFN (Germany)
- Simon Leinen, SWITCH (Switzerland)
- Guenther Schmittner, Universitaet Linz, ACONET (Austria)
- Jean-Marc Uze, RENATER (France)
- Celestino Tomas, REDIRIS (Spain)

## Dates and Phases

This subproject depended on a reliable signaling infrastructure, so the start of the experiments was delayed until the PNNI-setup was stable.

1. Research and planning: 97-07 to 98-01
2. Experiments: 98-02 to 98-03
3. Reporting: 98-03 to 98-04

## Network infrastructure

The NHRP tests used the SVC-service on the overlay network. This was a fully PNNI-based signaling system, as opposed to the tests in phase 1 of TF-TEN, where manual routing entries had to be entered in all the switches.

The NHRP network itself can be seen in [Figure 1](#). The ATM switches are excluded from the map. In all places, Cisco routers with a recent IOS were used. We used PVCs between the routers on the routed path because in a larger environment there will most likely always be some traffic to/from a router keeping an eventual SVC up. The routed path closely followed the underlying infrastructure that the SVCs had to traverse because we had no native signaling, only tunneled signaling. In a production environment given those conditions, this would be the optimal way to configure it.

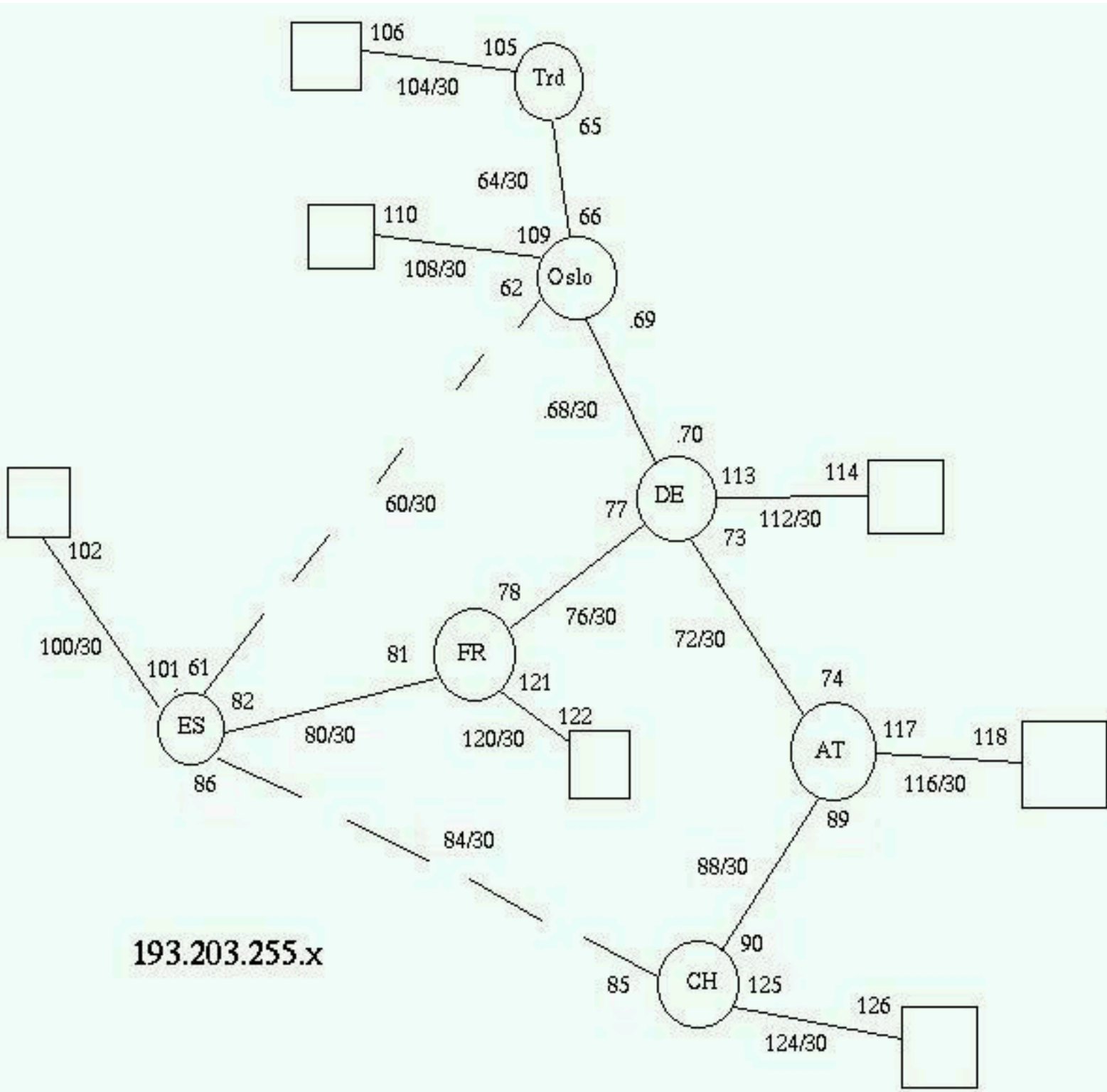


Figure 1: Network infrastructure. Dotted lines means interfaces were added at a later stage.

## Results and findings

### Background

In a backbone-network using routers connected to the same ATM-network, it is undesirable that traffic

have to traverse several routers on the way to the destination. However, for shortcuts to be set up through the network, one has to be able to exchange and look up ATM-level addresses. NHRP is one such mechanism that could work on a backbone level in an european backbone, and perhaps in national academic networks.

## **Current NHRP implementations**

There are not too many NHRP implementations available yet. Cisco has had its implementation fairly long, and IBM has an implementation. I have yet to see any other offerings in the area. However, NHRP has not yet become a standard, so this is not too strange.

The Cisco NHRP implementation is fairly stable, and we experienced no instability problems with it.

## **Phase One**

First, we configured the NHRP network with static routing. There was only one serious problem, namely that traffic from the workstations did not cause shortcuts to be set up before some time had passed. We were not able to resolve this problem due to limited time and a lot of activity going on at the same time. However, the mechanisms worked very well when the traffic originated in the router. The problem might be just a configuration-issue which we did not manage to look into.

## **Phase Two**

After having implemented NHRP with static routing we added a routing protocol(OSPF) on top of the network, and removed the static routing entries. As expected, NHRP worked exactly as before.

Then we added some redundant links. The new links were discovered and used by NHRP. We did not experience any problems because of this. NHRP has loop prevention mechanisms built in, and at least in one occasion, this protocol stopped an NHRP set-up-loop.

Due to lack of time we did not manage to do more advanced tests. It would certainly be interesting to purposely add routing loops to see how well NHRP loop detection works.

## **Observations**

The difference in round-trip between routed packets and packets sent over a shortcut were not large in general; only between 2 and 6 milliseconds, depending on the number of saved router hops. This was with an NHRP topology that followed the ATM topology in such a way that SVCs generally would have to go the same way. With native signaling we would most likely have seen more improvement, depending on the possibilities to route different ways in the ATM-providers network.

However, when the packets had to traverse routers, there were occasionally packets with a significantly

larger round-trip time. These will increase in number on a loaded router. In one case, the routed path was actually quicker than the shortcut path. This was due to the shortcut not going the quickest of two possible routes. The reason for this is a PNNI issue, not an NHRP issue, thus we did not investigate it further.

We saw occasional packet-loss due to shortcuts not being properly set up yet, after the SVC setup message for the shortcut was sent. However, in a well-configured network one should be able to choose to forward all packets over the routed path until the shortcut is set up. This was not working properly at all places/times in our network.

Setup-times are longer than for signaling, because one will first have to request and wait for the NBMA address. Typical time for this task is around 100-300 milliseconds in our network. Because the PNNI infrastructure we relied on was an experiment in itself and was down over large periods and because of equipment that broke, we were not able to collect real statistics on this. Additional setup-time is the actual SVC setup time, and this is investigated properly by another experiment.

## Discussions

### What NHRP provides

When NHRP is used there is some overhead with exchanging NBMA addresses. However, this overhead should be outweighed by the amount of data that bypasses the routers. Thus the potential for savings on router load is large. The switches will also benefit from a shortcut because once a shortcut is set up the packets do not have to traverse the switch twice, as is the case if it were to go into a router. This is at the expense of larger switching tables.

NHRP is not a routing protocol, it is merely a supplement. Thus, it will often be used in conjunction with protocols such as OSPF, which we used in our case.

### Description of the protocol basics

The components in NHRP are NHRP Clients (NHCs) and NHRP Servers (NHSs). The last ones will usually reside in routers, while NHCs can also reside in workstations.

A NHC keeps a local cache of IP-to-NBMA (for example ATM-address) mappings. It will use the service of one or more NHSs to resolve these. The NHC is the endpoint for shortcut connections.

A NHS will forward NHRP messages from NHCs, and will also send NHRP messages itself. The decision to set up a shortcut depends on the traffic the NHS has to send to NHSes that is in the same NHRP domain. Thus, NHRP can work without separate NHCs.

A router that detects traffic towards a destination reachable through a router that is on the same NHRP

network may choose to send a NHRP Resolution Request to try to find a more direct path through the ATM network. That next router might either send it further or give a reply, depending on whether the destination is reachable via other routers on the NHRP network or is served by the router itself.

If the router has its ATM address in the cache or serves the destination it will return an NHRP Resolution Reply. The original router will then set up a shortcut, and subsequent traffic will use that shortcut. Parameters in the request say whether cached info can be used or the resolution reply will have to travel all the way to the router that serves the destination.

If a router does not know the ATM address of the destination and do not have any routers to forward the NHRP Resolution Request to, it will send a negative NHRP Resolution Reply.

If the router has to forward the request to another router on the same NHRP network, that router gets the same choice of either forwarding or replying.

Other protocol messages that take care of cache invalidation and error reporting will travel on the network in the same fashion.

## Scalability issues

- Client level: With large amounts of traffic there is the possibility that the NHRP caches will grow to use a large part of the memory. In addition, mainly in the case where NHRP is extended all the way to the desktop and not only between the routers, it is possible that the shortcuts will become too numerous. Some ATM interface cards have limitations on the number of VCs per interface.
- Logical IP Subnet(LIS) level: The number of NHRP requests might become too large for a single NHRP server (NHS) as the NHRP domain grows. To compensate for this, it is suggested to use Server Cache Synchronization Protocol(SCSP) to maintain multiple synchronized NHSs within the LIS. This will reduce this problem.
- Domain level: For NHRP to work properly, every or most routers in the routing domain have to understand NHRP. The scalability of NHRP itself is about the same as the scalability of the routing protocol used in conjunction with it. Thus, the important thing to understand is that a non-NHRP-aware router on the path breaks the possibility to set up shortcuts to and beyond it.

## Limitations

- NHRP does not carry any QoS/CoS information in itself. There are, however, recent proposals to extend the protocol with such features..
- NHRP requests do not cross borders between NBMA subnetworks.

## Other related protocols

### LANE

LANE is an ATM-forum standard for LAN emulation on ATM-networks. It emulates all LAN functions, like unicast, multicast and broadcast traffic, and supports all internetwork layer protocols. LANE can easily be integrated with existing LANs, and looks like a normal LAN to existing LAN applications.

The components in an emulated LAN is Lan Emulation (LE) clients (LECs), LE servers (LES), Broadcast and Unknown Server (BUS) and LE Configuration Server (LECS), which relate to each other in the following way:

- The LEC performs data forwarding and address resolution, and provides a MAC level emulated Ethernet/IEEE 802.3 or IEEE 802.5 service interface to higher level software.
- The LES provides a facility for registering and resolving MAC addresses and/or route descriptors to ATM addresses.
- The BUS handle broadcast and multicast data, and also initial unicast data that is sent before direct connections are established.
- The LECS provides configuration information to clients that joins the emulated LAN.

However, LANE does not span multiple LISs, so it's not an alternative to NHRP, more a supplement.

### **Classical IP over ATM**

Classical IP over ATM supports only IP and only unicast. It simply provides an ARP-server, which the clients use to translate IP-addresses to ATM-addresses. This is a viable approach if one wants exactly this; direct IP-connectivity over ATM with other ATM-hosts in the same logical subnet.

The client will configure an ARP-server for the ATM interface, identified by its ATM-address. The client will, upon need of resolving the ATM-address corresponding to an IP-address on the same logical subnet, set up a VC to the ARP server(if it does not already exist) and ask for the ATM-address. The client itself is responsible for requesting the direct VC, the ARP-server merely provides the ATM-address.

Machines outside the Local IP Subnetwork (LIS) reachable via the same ATM-interface can only be accessed via a router.

Classical IP over ATM has thus limited features, however it's a simplistic protocol that does what it's supposed to do, and does it well.

### **Multiprotocol Over ATM(MPOA)**

Multiprotocol over ATM is a mechanism that combines LANE 2.0 with NHRP, allowing for shortcuts from workstation to workstation, using LANE with extensions in the local environment and NHRP between logical IP subnets. This protocol is said to scale to the corporate level, and would be worth a further experiment.

MPOA enables connections between different LANE subnets without requiring routers in the data path. However, routers will be used before direct shortcuts are set up.

MPOA clients(MPCs) are the edge devices in MPOA. MPCs will detect a flow of packets forwarded over an ELAN to a router with an MPOA Server (MPS). It may then ask the MPS for information to enable it to set up a shortcut to another MPC closer to the destination.

The ingress MPC will strip DLL encapsulation before forwarding it over the shortcut VC, and the egress MPC will again add it, so that it looks to the receiving destination as if it's travelled the normal routed path. To do this, the MPC will have to get information from the MPS to fill into a new DLL header.

MPSs use standard NHRP to exchange address information. NHRP is the only protocol used between MPSs, thus it's possible for two emulated LANs far away to be tied together using MPOA, as long as there's an NHRP network in between.

MPOA is a combination between LANE and NHRP, and certainly promises some interesting features. However, there's no MPOA implementations yet, at least not that we were able to find, so although it was a goal of this experiments, we were not able to test it.

### **Multicast Address Resolution Server(MARS)**

MARS is used to emulate multicast (and broadcast) functionality on ATM. There are two possible ways MARS can operate; it can resolve a multicast (or broadcast) address to a set of ATM addresses, or it can resolve it to the ATM address of a Multicast Server (MCS), which is responsible for forwarding the multicast or broadcast traffic to the final destinations.

ATM point-to-multipoint connections are used to transfer data to the end-systems. With leaf-initiated joins in UNI 4.0, MARS should be reviewed.

## **Conclusions**

NHRP can work well on an european ATM-based backbone network. Without NHRP, it would become tedious and hard to maintain a full mesh of connections between routers, which is one of the advantages you would probably want to use in ATM. NHRP provides a way to add links dynamically as the need comes.

Due to lack of support from the underlying networks, we were not able to test anything but UBR SVCs over CBR tunnels. Thus, our tests were merely tests of how NHRP itself worked, not real traffic tests. This means that we can not conclude anything about the reduced amount of traffic that would have to pass through routers. This will also depend on the nature of the traffic and whether the traffic will last or is just a short burst.



## References

1. James V. Luciani, Dave Katz, David Piscitello, Bruce Cole, Naganand Doraswamy: NBMA Next Hop Resolution Protocol(NHRP). Work in progress.
2. Classical IP and ARP over ATM, RFC1577, January 1994
3. LANE v2.0 LUNI Interface, ATM Forum, July 1997
4. Multi-Protocol Over ATM Specification v1.0, ATM Forum, July 1997
5. Signalling, PNNI and RSVP tests for TF-ten.
6. James V. Luciani, Grenville Armitage, Joel Halpern, Naganand Doraswamy: Server Cache Synchronization Protocol (SCSP). Work in progress.
7. James V. Luciani: A distributed NHRP Service using SCSP. Work in progress.
8. Multicast Address Resolution Server(MARS), RFC2022, November 1996.
9. UNI 3.1, ATM Forum, 1994
10. UNI 4.0, ATM Forum, July 1996

## 4.10. ATM Addressing

### Experiment Leader

Kevin Meynell, TERENA

### Summary

ATM addressing is an area that appears not to have been investigated in depth by network providers, yet it is essential to the operation of ATM networks using signalling. Public Telecommunications Operators (PTOs) have indicated a preference for E.164 addresses that will allow legacy networks to be migrated to ATM without having to undertake major re-numbering. Such addresses however, are expensive, scarce and inflexible and many National Research Networks have decided to adopt NSAP-based addressing schemes. Unfortunately, whilst address formats are well-defined, there are still no standards for translating between the different schemes. The consequences of integrating research and public networks are still unclear.

### Participants

TERENA

### Results and Findings

This aim of this project was to investigate the issues relating to ATM addressing and produce a recommended ATM addressing scheme. ATM addressing is an area that is not well understood and poorly documented, yet it is essential to the operation of next-generation ATM networks. This paper discusses the various issues, and summarises the views of the European research networking community.

ATM is a connection oriented technology. This means that a virtual circuit needs to be established across an ATM network prior to any data transfer. There are two types of virtual circuit: Virtual Paths (VPs), and Virtual Channels (VCs). A VP is a bundle of VCs (usually 256) that are switched transparently across a network with a common Virtual Path Identifier (VPI). An ATM switch receives a cell with a particular Virtual Channel Identifier (VCI), and then uses a local translation table to determine the outgoing port and the new VPI/VCI value of the onward connection.

There are two methods by which VCs are established. Permanent Virtual Channels (PVCs) need to be manually configured and are therefore cumbersome to manage. Until recently however, these were the only type of VC available. Switched Virtual Channels (SVCs) are set-up automatically by signalling protocols (similar to how telephone calls are made), making them more flexible and easy to manage. Unfortunately, SVCs have only recently become available which means that most ATM networks currently use PVCs. Nevertheless, SVCs may eventually replace PVCs entirely.

Several types of signalling protocol have been defined (e.g. UNI 3.0/3.1/4.0), but whatever protocol is used, an addressing scheme is required to allow the source and destination of connections to be identified. To this end, the [ATM Forum](#) has defined two types of addressing scheme that may be used for ATM networks.

The [International Telecommunications Union \(ITU-T\)](#) decided that E.164 addresses (telephone numbers) should be used for public ATM networks. This allows legacy PTO networks to be migrated to ATM without having to undertake major re-numbering. An E.164 number is up to fifteen digits in length (including a three digit international prefix), with an associated twenty octet sub-address field. The sub-address field provides additional addressing capacity outside the E.164 numbering plan, and is handled by equipment at subscriber (end-user) premises.

Unfortunately, whilst the E.164 recommendation theoretically provides a large address space, much of it has been allocated inefficiently. The geographically-based numbering plan dates back to mechanical telephone exchanges when technical considerations prevented efficient number allocation. These problems have been accentuated by the dramatic growth in devices connected to public telecommunications systems. This is demonstrated by the constant changes in area codes, and partial renumbering of subscriber addresses. Furthermore, E.164 addresses can only be obtained (by subscribers) from telecommunications companies, and are an expensive resource.

As a consequence, the [ATM Forum](#) defined an address format for private ATM networks, such as a National Research Network (NRN). These are based on OSI Network Service Access Point (NSAP) addresses that were originally defined for X.25 networks. Strictly speaking, a private ATM address is not an NSAP, but this term has passed into common usage.

An NSAP address is twenty octets in length (the term octet is used as byte size can vary between different platforms) and consists of three components. The Authority and Format Identifier (AFI) identifies the type of NSAP, the Initial Domain Identifier (IDI) identifies the address allocation and administration authority, whilst the Domain Specific Part (DSP) contains routing information.

There are three types of NSAP that may be used for ATM networks:

**E.164 Format:** The IDI is an E.164 number. This allows E.164 addresses to be used within private ATM networks. The address prefix is based on an E.164 number, with individual nodes being identified by the remaining octets. The AFI value is 45.

```

+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
| A |           |           |           | S |
| F |      E.164 | HO-DSP |      ESI | E |
| I |           |           |           | L |
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+

```



of PTO monopolies), whereas modern networks often operate internationally and compete with others in the same geographical area. This means it is often necessary to assign addresses logically as well as geographically.

Another advantage of NSAP addresses, is that a registry system (a hierarchy of organisations that allocate addresses) is already in operation. These generally provide addresses at a much lower cost than PTOs (who are forced to make E.164 addresses expensive to limit demand) which is especially important to NRNs.

For these reasons, most NRNs have indicated they would prefer to use NSAP addresses. Whilst this is not a problem at the present time (as the few NRNs using switched services are doing so over tunnelled connections), it will become an issue if PTOs start using switched services for their core networks.

In theory, the interconnection of public and private networks operating different address schemes should not be a problem. [ATM Forum](#) standards state that where a call originates from, and is destined for, networks supporting NSAP addresses, the egress switch on the originating network will move the destination NSAP address into the E.164 sub-address field. This will be replaced by the E.164 address of the switch providing ingress to the destination network, allowing the call to be forwarded across the public network. Once the call reaches the ingress switch, the NSAP address will be moved back into the main address field.

Where a call originates from a network supporting NSAP addresses and is destined for a network only supporting E.164 address, the destination address should be coded as an E.164 format NSAP. The egress switch will then convert this to a native E.164 address, allowing the call to be forwarded to its destination.

Finally, where a call originates from a network only supporting E.164 addresses and is destined for a network supporting NSAP address, the destination NSAP address should be encoded in the E.164 sub-address field. The main address field will contain the E.164 address of switch providing ingress to the destination network. Once the call reaches the ingress switch, the NSAP address will be moved into the main address field.

Unfortunately, it is not clear how network switches obtain the information to map NSAP addresses to E.164 addresses and vice versa. It seems likely that some manual configuration will be necessary in the short-term, although there are proposals to establish a public directory service. At the time of writing, address translation implementations had become available for a few ATM switches (e.g. Cisco LightStreams and Fore ASX-200s), but these were not released in sufficient time to be tested by the TF-TEN group.

Another potential problem is that some ATM switches used by PTOs are rumoured not to fully support the E.164 sub-address field. This would have obvious consequences for NSAP-based networks interconnected by an E.164-based network.

European NRNs are increasingly using ATM to provide their backbone networks. Whilst none of these ATM networks use signalling for their production services at the present time, they are likely to be amongst the first users taking advantage of SVCs. Nevertheless, most (if not all) NRNs purchase their connectivity from PTOs, and are therefore dependent on what they choose to provide.

For reasons outlined in this paper, NSAP addressing schemes are preferable to NRNs. Indeed, [JANET](#) (UK), RENATER (France), [SURFnet](#) (Netherlands) and [Uninett](#) (Norway) have already drafted such schemes. Therefore, the position of the academic networking community is that PTOs must offer NSAP-based signalling to NRNs. If address translation can be successfully implemented by PTOs, then of course it is not an issue whether they use E.164 addresses for their core network. Nonetheless, the onus should not be on the NRNs to provide this themselves. In fact, nor could they unless the PTOs revealed the structure of their networks; something they have always considered unacceptable.

## Further studies

The testing of address translation implementations should be considered a priority. It is also necessary to continue to monitor standards relating to addressing from the ATM Forum and the ITU.

## Bibliography and references

1. [ATM Addressing Discussion Paper](#)
2. [ATM Internetworking](#)
3. [JANET ATM Addressing Scheme](#)
4. [UNI Signalling Specification 3.1](#)
5. [UNI Signalling Specification 4.0](#)

## 4.11. Performance of the Native ATM Protocol and native ATM applications.

### Experiment Leader

Stefania Alborghetti, INFN, Milano

### Summary of results

From the initial survey it was clear that native ATM implementations are necessarily bound to the underlying API. Due to a lack of standard ATM APIs, every particular native ATM implementation seems to rely on a particular ATM adapter and it is therefore very difficult to find a native ATM application capable of being implemented in a multi-vendor environment.

We concentrated on public domain software. Two interesting implementations were considered:

- *Tcp and Udp Over Non Existent IP (Onip)*, maintained by *Enst, France* (see participants below).
- *Arequipa*, maintained by *LRC (Laboratoire de Reseaux de Communication)*, *Switzerland* (<http://lrcwww.epfl.ch/arequipa>).

We were able to implement and test the *Tcp* and *Udp Onip* on the wide area, while unfortunately the *Arequipa* implementation could not be tested. Nevertheless it was carefully studied and a description is reported.

The *Onip* stack was successfully tested with two applications: an *http* client and an *mpeg* viewer. The performance was evaluated in terms of packet inter-arrival times. The purpose was to verify that the sending workstation was transmitting at a rate determined by a traffic algorithm applied at the source. Such algorithm is implemented in the kernel modules and its purpose is to improve significantly the performance of *tcp* and *udp* layers over an *atm* network by preventing the sending station to violate the *UPC* contract in the network. Such algorithm is therefore based on the *gcr* mechanism and accepts *UBR*, *CBR* and *VBR* parameters. Tests were performed on *PVCs* with the *UBR*, *CBR* and *VBR* classes.

As far as native ATM APIs are concerned, standards exist but are currently not extensively implemented. To our knowledge two documents are available:

- a semantic description by the *Atm Forum* [1];
- the *X/Open CAE Specifications* by the *XOpen Group* [2].

The performance of the *Fore SPANS* API was tested with the *netperf* utility on the local area. This API is included in *ForeThought* software and is based on the proprietary *Fore* signaling protocol *SPANS*. Since it is not *UNI* compliant, it was not tested any further in the wide area.

Fore has recently released a new API, which is UNI compliant, and which is based on the XOpen specifications. This API also incorporates the Winsock2 specs. and is thus available for Windows95 and NT architectures in addition to the Sun and Sgi architectures.

## Participants

*INFN* (Italy) , *Enst\** (France), and *Cnet\** (France).

(\*) Not part of the task force TEN.

## Dates and Phases

- Survey of existing native atm implementations: June-September 1997
- Performance of the Fore SPANS API: October 1997
- Performance of the Tcp and Udp Over Non Existent IP implementation: February 1998

## Network infrastructure

### 1. Performance tests: Fore SPANS API.

Tests were performed between a Sun Sparc 20 workstation running Solaris2.5 and a Silicon Indy running IRIX 5.3. The sun station was equipped with a Fore sba200 adapter while a Fore gia200 adapter was present on the SGI station. The two workstations were connected at 155 Mbits/s through a Fore ASX-200 Switch.

### 2. Performance tests: Tcp and Udp over non existing Ip.

Tests were performed between two Sun workstations at INFN, Italy, and at CNET, France. Both workstations were equipped with a 155 SunATM SBus adapter. In Italy the workstation was connected to a Cisco Lightstream 1010 while in France a Fore ASX-200BX was used. The two switches were connected by a 2 Mbits/s VP through the James network. Shaping was enabled on both switches.

## Results and findings

### 1. Initial survey: standards, APIs, applications.

*Standards.*

As reported above in the summary section, standard API are not extensively deployed yet. We were able to have access to the following two documents:



- A document by the ATM Forum which specifies a semantic definition of native ATM services, advancing therefore the development of APIs. Such document addresses concerns with interoperability aspects and proper abstraction of ATM procedures and parameters. See reference [1] for more details.
- A document which is a CAE (Common Applications Environment) Specification by the X/Open group (<http://www.rdg.opengroup.org/>). This document describes the industry-standard open system interfaces to communication services. These include two APIs: Sockets and XTI. The ATM transport protocol is outlined as a preliminary specification both for Sockets and XTI. The following are supported:
  - a subset of the functions specified in the UNI version 3.0 and 3.1;
  - AAL5 messages.

See reference [2] for more details.

### *Interesting APIs.*

Among the existing ATM APIs three are particularly interesting:

- The Windows socket 2.0 (winsock2.0) by Microsoft, which includes ATM specific extensions. The winsock2.0 supports:
  - UNI 3.0 and 3.1 functions;
  - AAL5, plus a user defined AAL messages.

The plans are to incorporate UNI 4.0 and other AALs in future releases. See reference [3] for more details.

- The linux ATM API, which is part of the public domain ATM distribution for Linux. The following are supported:
  - UNI 3.0 and 3.1
  - AAL5 plus raw "AAL0" messages

See reference [4] for more details.

- A Fore API, which is UNI compliant and incorporates both the XOpen Specs and the Winsock2 specs. It was announced by Fore in October 1996 and is part of the new versions of the ForeThought software. The first product available was a ForeThought Winsock 2.0 ATM API. A second set of ForeThought APIs was based on the X/Open XTI specifications. This API makes possible to develop native ATM applications for Windows

95/NT and for UNIX environments. Being based on the X/Open XTI specifications, it is expected to support what has been stated in the previous section about standards.

### *Arequipa.*

Arequipa (Application REquested IP over ATM) was developed by W. Almsberger, J. Le Boudec and P. Oechslin [6] and is part of the standard ATM distribution for Linux [4]. There is also an RFC available [8]. The goal of the Arequipa implementation is to provide the ATM Quality of Service to TCP/IP applications, assuming end-to-end connectivity. Arequipa does not require changes in the network but changes are required:

- in the TCP/IP protocol stack of the end stations, particularly at the socket layer;
- in the code of the applications in order to make use of the extension of the socket interface. This extension consists of only four calls for setting up, tearing down and modifying the traffic parameters of connections.

Traditional TCP/IP applications can then request a connection with specified traffic parameters. Arequipa uses the signaling daemon to establish SVCs with the parameters requested by the application itself. All the data traffic relative to the application will then be forwarded on this newly created VC. To achieve this the modifications added to the socket layer permit to route all the traffic relative to the application to a virtual driver called Arequipa. In particular the route cache entry in the socket descriptor (which point to the interface to which the data have to be forwarded) is set to point to the Arequipa driver. Besides a further entry containing a pointer to the vc created is added to the socket descriptor. The Arequipa driver simply sends the data to the vc indicated in the socket descriptor.

Arequipa is not a native ATM implementation because the standard TCP/IP stack is still used. Nevertheless applications gain access to the ATM Quality of Service features.

So far the following applications have been modified to use Arequipa:

- the Arena browser and the cern httpd web server [6];
- the VIC video conferencing tool [9].

The two following new important functionalities have been recently added to the Arequipa implementation:

- UNI4.0 signaling;
- Q.2963.1 connection modification capabilities.

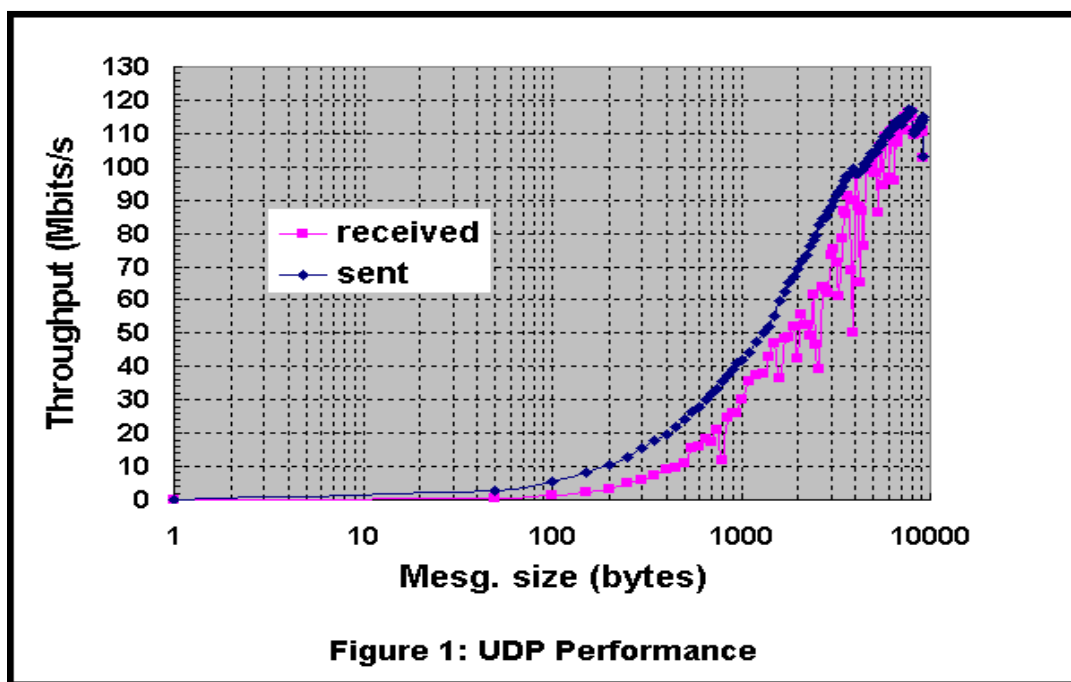
Therefore the Arequipa-capable version of VIC is able to negotiate and renegotiate CBR connections at run time specifying different PCR values. This is certainly an important task that has been achieved.

## Tcp and Udp Over Non Existing IP.

Tcp and Udp Over Non Existent Ip (ONIP) is a stack available for the Solaris environment that offers ATM services to the Internet applications . It provides the usual socket and TLI interfaces so that applications need just minimal modifications in order to run on top of it. Domain names can be resolved into IP addresses leading the the normal Tcp/Udp-IP stack, or into ATM addresses leading to the use of the Onip stack. Socket or TLI connections can in fact be mapped directly on Switched virtual circuits. Nevertheless permanent virtual circuits can be used as well. Quality of Service is thus provided to applications both on SVCs and PVCs. It is possible to multiplex data on a single VC, like in classical IP, or to have one guaranteed VC per application. A flow control mechanism based on the *gcr* algorithm is implemented within the Solaris kernel in the Tcp and Udp layers. The purpose is to apply an UPC contract on a VC basis, something that is not done by standard ATM adapters since it would require a fair multiplexer implemented on the hardware. QoS requirements are usually specified i by the client and communicated to the server in the set-up phase. See subsection 3 for more details.

## 2. Performance tests: Fore SPANS API.

The Fore SPANS API is part of the ForeThought software and allows developers to access directly the ATM services. Since the netperf benchmark utility includes a section based on this API, it was possible to measure the throughput between two workstations. The netperf utility was developed by Rick Jones at HP. The two workstations were connected at 155 Mbits/s. The switch between them was a Fore ASX. therefore the SPANS signaling protocol could be used. The main idea was to compare the results of the



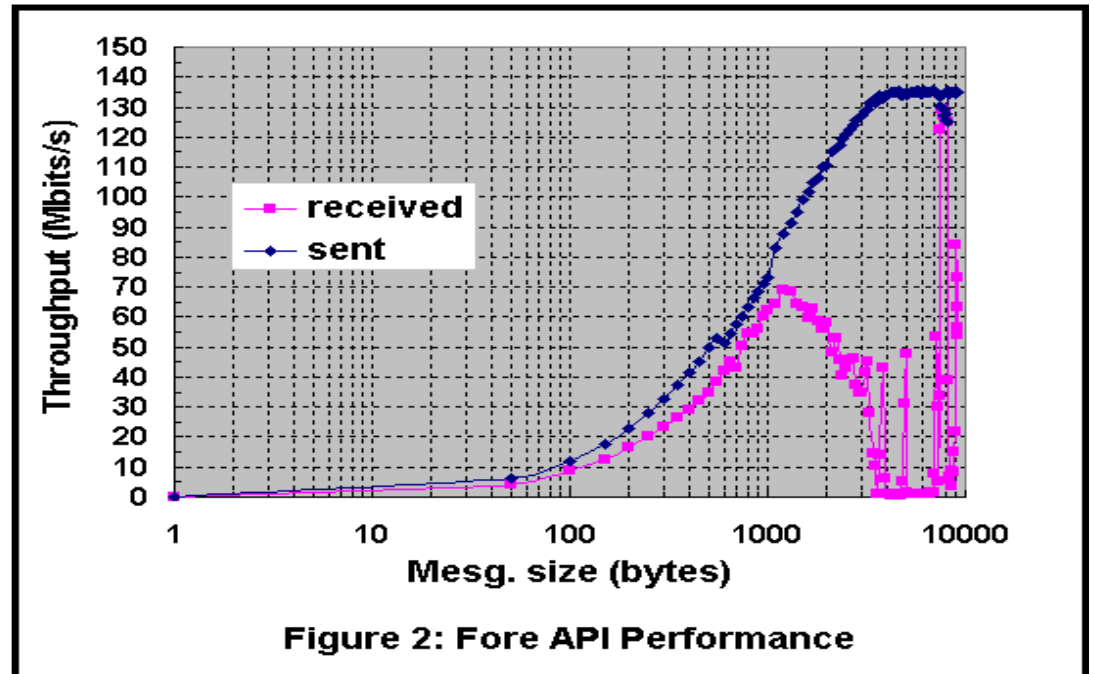
netperf Fore test with the results of the netperf UDP tests. The two protocols are in fact both unreliable and work in the same way, i. e. flow control and reliability are left to the upper applications. In this particular case, netperf tests, none of them was implemented.

Figure 1 shows the results of the UDP test while figure 2 shows the results of the tests performed with the Fore

API. In both figures the throughput (in Mbits/s) is plotted as a function of the message size (in bytes). The blue curve represents the sent throughput while the pink curve represents the received throughput as

communicated by the receiving station at the end of the netperf test. Tests were performed at night time with the two workstations completely unloaded and dedicated to the tests. Figure 1 shows that in the case of udp tests, the throughput increased as expected. Nevertheless the maximum value that could be reached was 117.06 Mbits/s, while the theoretical value is 134.69 Mbits/s [14].

Thus roughly 13% of the bandwidth could not be used. We presume this was due to a poor performance of the sending workstation. The CPU load was in fact constantly 100%. Figure 2 shows that in the case of native tests the sent



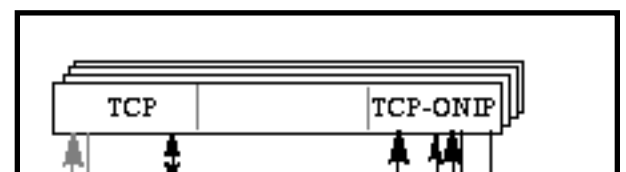
of native tests the sent throughput was higher than the one obtained in the udp tests. A value of 135.34 Mbit/s was reached while the theoretical one was 135.63 [14]. This improvement of the sending performance is not due to the overhead of the Ip and Udp protocols (in fact the two theoretical values differ only of 0.94 Mbits/s [14]) but to some other feature internal to the kernel of the sending workstation. On the other hand a very poor performance in the received throughput was observed when reaching values of roughly 70 Mbit/s. We currently do not know the reason of such a poor behavior.

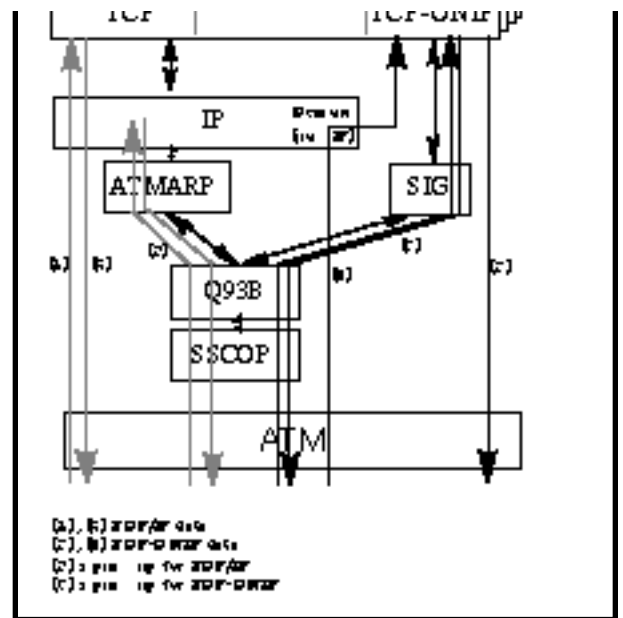
From the tests performed it could be concluded that:

- it was not possible to transfer data using the Fore SPANS API at values higher than 70Mbits/s;
- the performance of the native protocol is higher when transmitting data. Comparing the two maximum values that were reached (135.34 for Fore tests and 117.06 for Udp tests), the improvement is of roughly 13%.

### 3. Performance tests: Tcp and Udp over non existing Ip.

*Tcp/Udp-Onip architecture.*





The scheme here aside describes the protocol blocks that constitute both the classical IP stack and the Tcp Onip stack. A similar representation could be made for the Udp case. The classical IP implementation relies on an ATMARP module to handle IP-to-ATM address translation and connection management. With Tcp-Onip a module called SIG has been added. This module is responsible for connection establishment and release. It receives control messages from the upper TCP layer and manages the connection with the remote host using the ATM signaling module (Q.93B in the figure). The SIG module also deals with the case of permanent virtual circuits. An application chooses between the Tcp/Ip or the Onip stack by communicating the *sinaddr\_in* structure to the TCP layer. Two cases are possible:

- an IP address is present and the standard TCP/IP stack is used;
- the IP address is empty or equal to 0.0.0.1 and the Onip stack is used. In this case the TCP layer extracts the ATM address and the QoS parameters and communicates them to the SIG module. In case of SVCs the connection is then created and the SIG module returns the VPI/VCI values to the TCP layer so that data can be forwarded to the ATM layer. In case of PVCs (0.0.0.1) data is directly forwarded to the ATM layer using a standard VCI. Nevertheless QoS is guaranteed to the application since the shaping and scheduling algorithm is used.

To achieve its purposes the Onip stack contains:

- an *ATM name service* (ANS) capable of translating logical names into ATM addresses. Both NSAP and E.164 formats are supported. The interface to the ANS is similar to the standard DNS (*get\_hostbyname* function).
- a *packet scheduler* based on the *gcr* algorithm. This control mechanism was added in order to overcome to the performance issue arisen by the Tcp control mechanism. This latter in fact, not being aware of the UPC control performed at the UNI side of the network, might send non-conforming cells and lower the overall network performance. With a *gcr* flow control mechanism, implemented in the source by using the same parameters specified at the connection set-up, it can be only cells conforming to the

UPC are transmitted and this results in a better utilization of the ATM connection. To describe briefly how the control mechanism works, for each datagram received by the kernel from an upper application, an eligibility time is calculated according to a leaky bucket algorithm. The packets are then blocked until they are eligible to be sent by refusing them the access to the lower layers. For details see [11], [12] and [13].

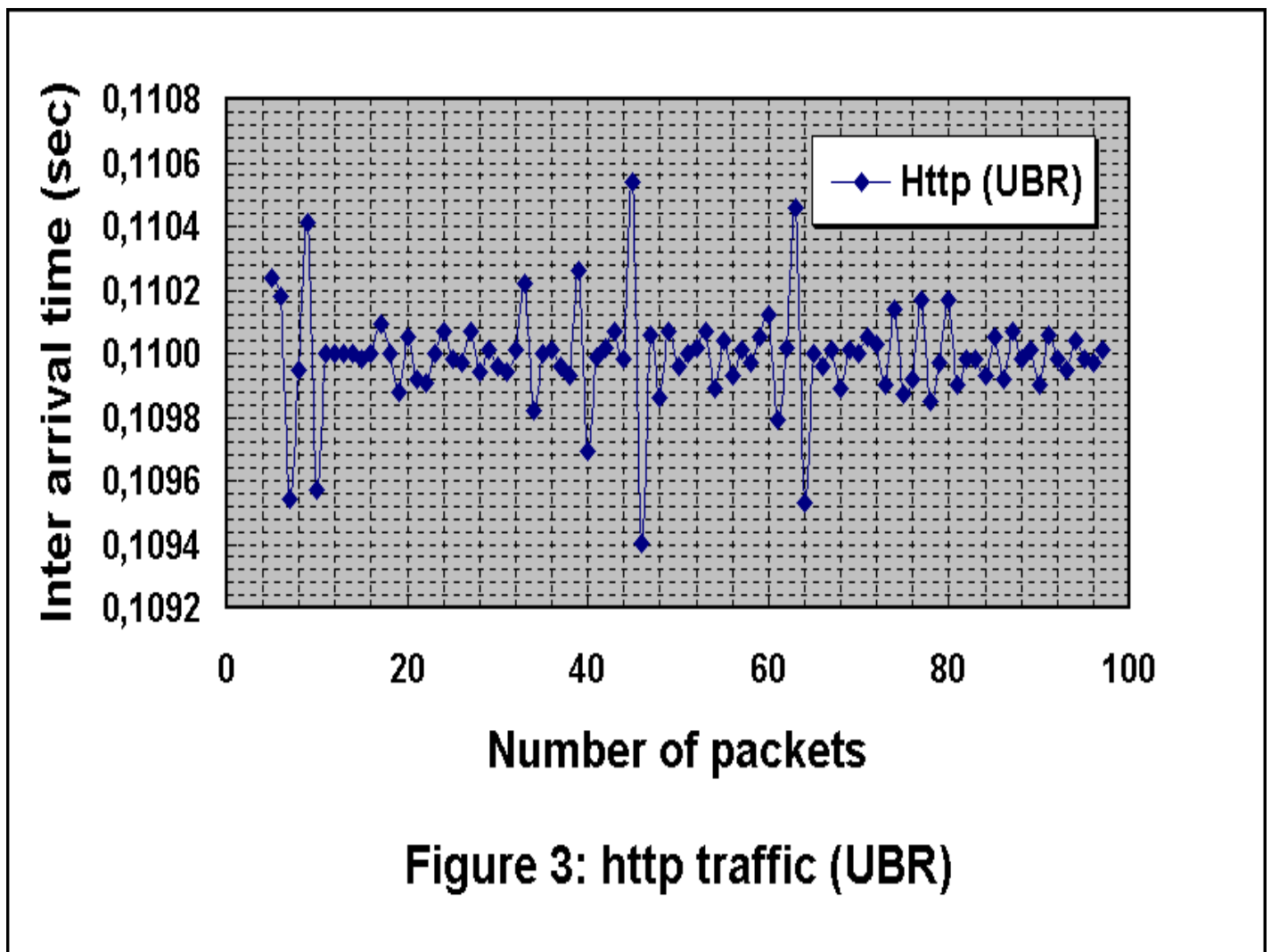
### *Tcp-Udp Onip performance tests.*

Tests were performed only on PVCs and not on SVCs due to major problems of interoperability between the signaling modules of the test bed. Nevertheless both http and video could be tested. The purpose was to analyze the performance of the shaping mechanism applied at the source. Such mechanism permits to specify different classes of services (UBR, CBR and VBR) even on the same VC. The http was tested with a Mosaic client modified to run on top of the Onip stack. The video was tested with an application called *MPX* that was listening on an UDP port. The client can request a particular film to the source which starts sending the requested film on the specified port. The MPX application waits on such port and displays the film as soon as it receives something. We took performance measures by analyzing the output of the *atmsnoop* command, which is included in the ATM distribution of SUN ATM SBus adapters.

We performed the following tests:

- http only, (UBR);
- video only, (VBR);
- video and http together, (video CBR and http UBR).

Figure 3 shows the inter-arrival time of packets of an http connection. The length of the packets was 8244 bytes, the total bytes transferred were 332003.



Although the PCR was set to the maximum (2.5 Mbits/s), the total bandwidth available on the WAN VP was of 2 Mbits/s. As it can be seen from the figure the average inter arrival time is of 0.11 seconds. Since the average size of each packet was 8244 bytes, we can assume for each packet 172 ( $8244/48+1$ ) cells were sent. Therefore an inter arrival time of 0.11 seconds corresponds to a bandwidth utilization of  $172/0.11 \sim 1500$  cells/s, which is much less than 2 Mbits/s ( $\sim 4770$  cells/s). We presume this lower value is due to the TCP protocol waiting for ack packets before continuing to send further data. By setting lower PCR values higher inter arrival times were observed.

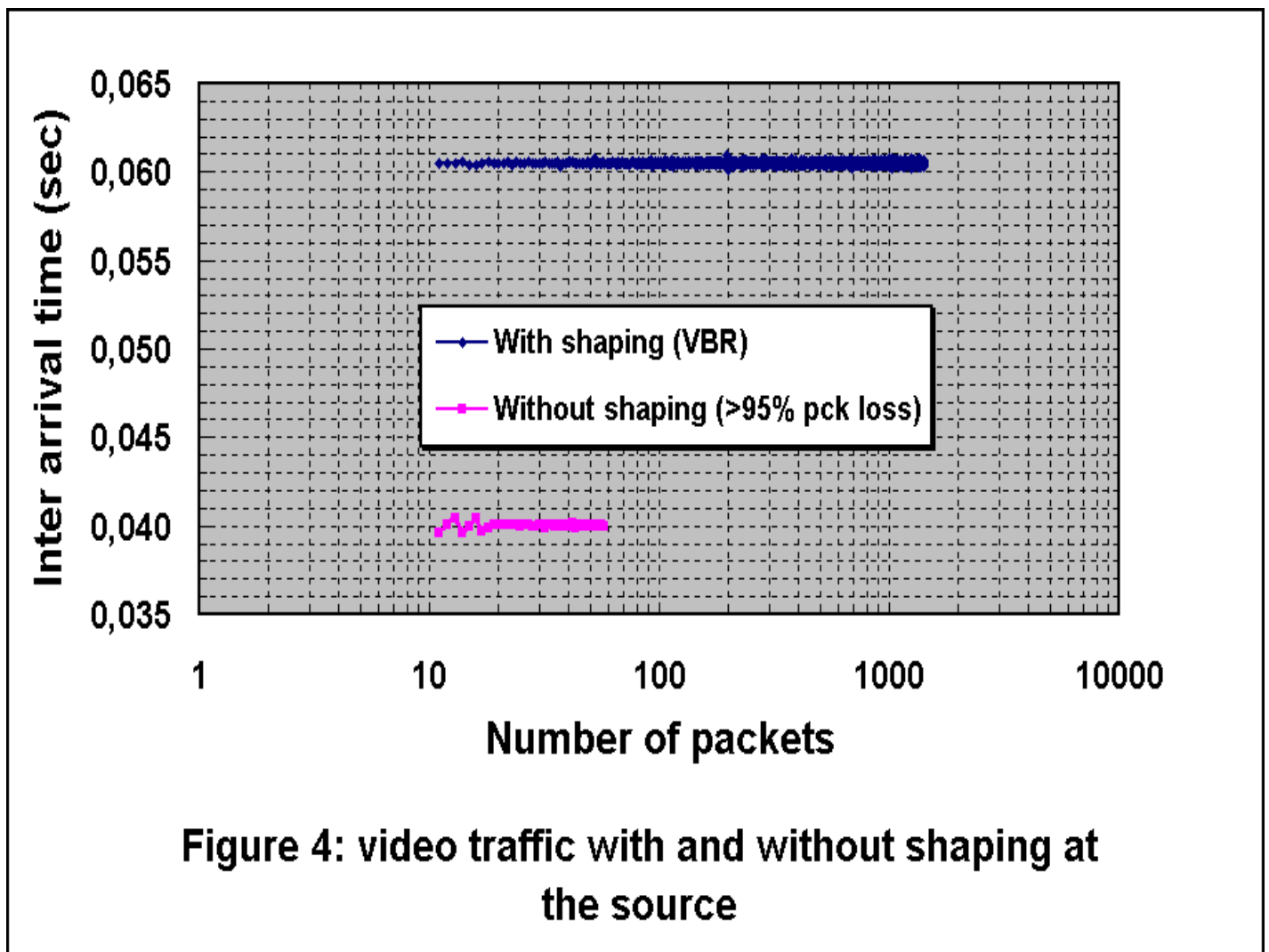
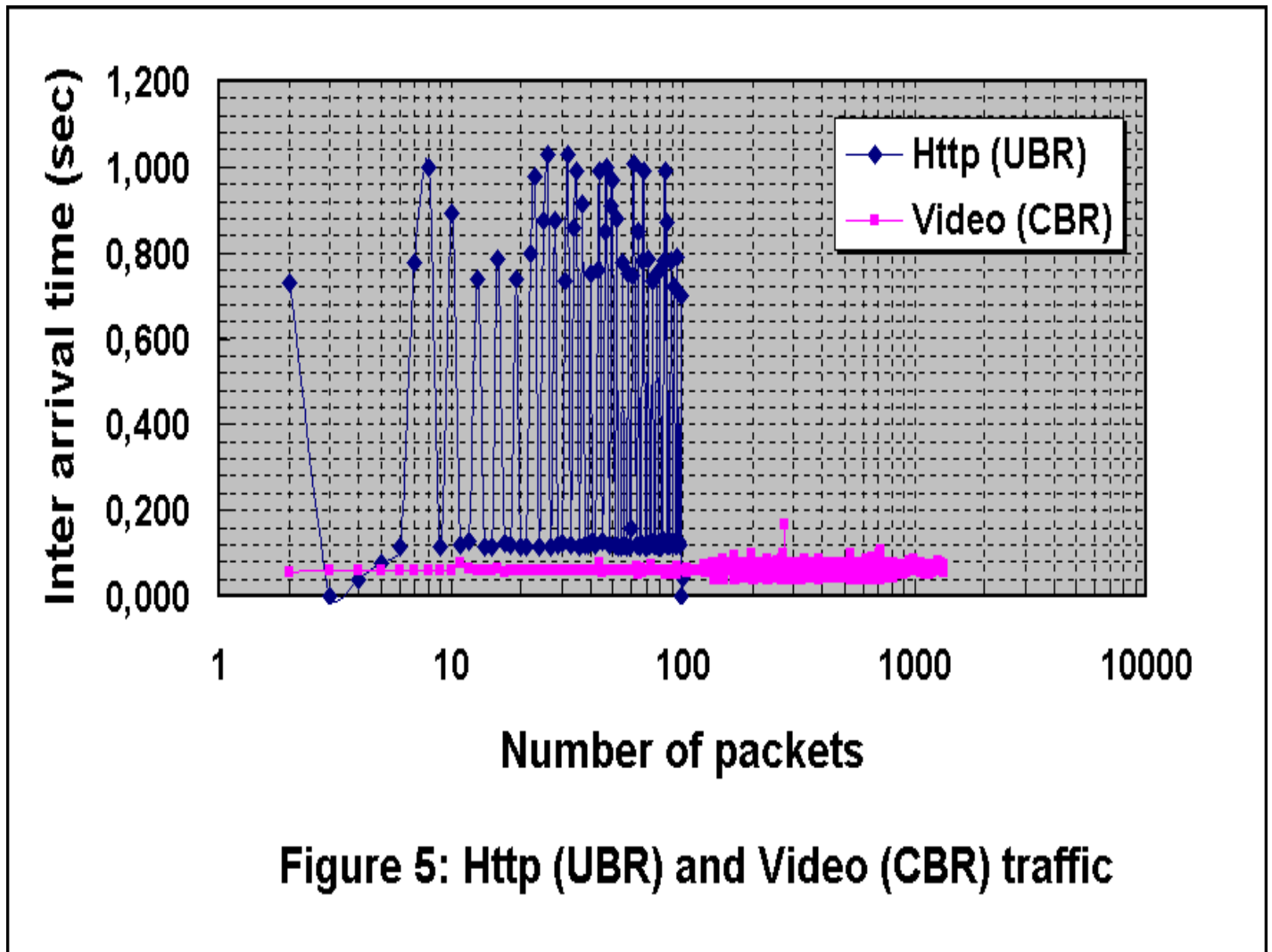


Figure 4 shows the inter-arrival times for video traffic. The upper curve represents the traffic relative to the transmission of an mpeg movie with VBR shaping applied at the transmitting source. The lower curve refers to the transmission of the same mpeg movie but with no shaping applied at the source. In this latter case, without shaping at the source, more than 95% of packets were lost and only a few frames of the mpeg movie could be seen. This was of course due to the fact that the source was sending too fast and therefore violating the network contract of 2 Mbits/s. It has to be noticed that even setting the speed of the atm card to a value of 2 mbits/s (*atmspeed* command) didn't improve the situation. On the other hand with shaping applied at the source the inter arrival time between packets was constant and no packets were lost. The mpeg movie could be seen with the same quality as when seen locally. The shaping algorithm applied at the source was based on the *gcr* algorithm. Both CBR and VBR parameters were used. In particular the PCR was set to 3000 cells/s. Considering that the length of each packet was of 9040 bytes (72320 bits) we expected to have a packet inter-arrival time of:

$$72320/(3000 \times 48 \times 8) = 0.063 \text{ secs}$$



The value that was obtained was 0.0605 secs. Figure 5 reports again the inter-arrival times between packets relative to an http connection (UBR) and a video connection (CBR).



**Figure 5: Http (UBR) and Video (CBR) traffic**

As it can be seen, since the video traffic is CBR the inter arrival time between packets remains constant to 0.06 while the inter-arrival time between UBR packets (http connection) fluctuates, in a considerable way, reaching values of even 1 second. Therefore it was still possible to see an mpeg movie with an extremely good quality even if an http connection was simultaneously running.

### Test related problems and general comments

It is important to note that if an application is to be made QoS aware, some code has to be added somewhere. Currently there are three approaches:

- plain native ATM, that is the application is modified entirely in order to substitute the standard

socket or XTI calls with the calls provided by a native ATM API. An example is the VIC tool which was modified in order to support the Fore SPANS API.

- applications can be made RSVP capable. Again this requires the code of the applications to be modified. Besides the network has also to be modified in order to support RSVP. An example is again VIC, which was also modified to support RSVP.
- modifications can be made in the TCP/IP protocol stack of the end-systems. This requires to modify both the TCP/IP stack and the application code but the modifications to the applications are usually minimal. The Onip and the Arequipa implementations are examples of this approach. The difference between them is that in the case of Arequipa the IP layer is still present while with Onip it isn't. Again VIC is an example of an application made QoS aware with this approach.

It is still to be understood which of these approaches is the best. It is our belief though, that rewriting an existing application from scratch in order to make it native might not be the cleverest thing to do. An approach like the third one might be the easiest in order to make all the already existing TCP/IP applications capable of exploiting the QoS ATM services. This has the clear advantage to be compatible with the case ATM is not entirely deployed on the network. In addition it could also provide a smoother transition towards purely ATM native applications.

## Acknowledgements

We would like to thank *Enst* and *Cnet* for their collaboration with the tests of the Onip stack and for their availability to explain the details of such implementation.

## References

[1] The Atm Forum Technical Committee: Native ATM Services: Semantic Description Version 1.0 (<ftp://ftp.atmforum.com/pub/approved-specs/>)

[2] X/Open CAE Specification XNS Issue 5: <http://www.opengroup.org/public/pubs/catalog/c523.htm>

[3] Winsock 2.0 home page: <http://www.intel.com/ial/winsoc2>

[4] ATM on Linux home page: <http://lrcwww.epfl.ch/linux-atm/>

[5] Arequipa home page: <http://lrcwww.epfl.ch/arequipa>

[6] Werner Almesberger, Jean-Yves Le Boudec, Philippe Oechslin: "Application Requested IP over ATM (Arequipa) and its use on in the web", Proceeding of the 3rd symposium on Inter-networking, Nara, Japan, October 1996 ([ftp://lrcwww.epfl.ch/arequipa/iw\\_full.ps](ftp://lrcwww.epfl.ch/arequipa/iw_full.ps)).

[7] Werner Almesberger: "Arequipa: Design and implementation", ATM Forum contribution to

the December '96 meeting in Vancouver ( [ftp://lrcwww.epfl.ch/arequipa/aq\\_di-1.tar.gz](ftp://lrcwww.epfl.ch/arequipa/aq_di-1.tar.gz) ).

[8] Werner Almesberger, Jean-Yves Le Boudec, Philippe Oechslin: RFC 2170, "Application REQuested IP over ATM (AREQUIPA)" .

[9] Werner Almesberger, Leena Chandran, Silvia Giordano, Jean-Yves Le Boudec, Rolf Schmid: "Using Quality of Service can be simple: Arequipa with Renegotiable ATM connections", submitted to Elsevier Preprint.

[10] Hossam Afifi, Dominique Bonjourn, Omar Elloumi: "TCP Over Non Existent IP for ATM Networks", JENC7 proceedings.

[11] Dominique Bonjourn, Omar Elloumi, Hossam Afifi: "Internet Applications over native ATM", to appear in Computer Networks and ISDN.

[12] Leila Lamti, Hossam Afifi, Manuel Hamdi: "Design and Implementation of a flexible traffic controller for ATM connections", proceedings of 7th IFIP Conference on High Performance Networking (HPN'97) pages 130-146 - New York, April 1997.

[13] Leila Lamti-98-1 Enst technical report, <ftp://rennes.enst-bretagne.fr/pub/incoming/Report.ps>.

[14] John David Cavanaugh: "Protocol Overhead in IP/ATM Networks", Minnesota Supercomputer Center, Inc.

## 4.12. ATM Network Management

### Experiment leader:

Zlatica Cekro, University of Brussels, VUB

### Summary of results

The experiments were based on the tests of functions for M1, M2 and M3 management views defined by the ATM Forum. Eleven NRN participated actively in this phase. Experience is gained in domain of management and system interoperability at the management plane functions of configuration, performance, fault and security. Special test results concern the OAM flows, Web based management, SNMP based MIB extensions for ATM and Xuser interface tests.

SNMP versions 1 and 2 (community based access control) were used. A Management Platform based on SunNet Manager-SunNet Domain Manager version 2.3 on Solaris 2.4 was located at the University of Brussels. The Management Platform enabled "monitoring" i.e. access and read only class of management service for eleven NRN ATM switches and three routers with an ATM interface. Beside this, the "configuration" i.e. write class of management service for two NRN ATM switches was enabled.

The transport links between the Management Platform and NRN ATM devices were realized by the operational Internet service and by the ATM Permanent Virtual Connections (PVC) configured over the TF-TEN Overlay network. The test results concern the following:

The SNMPv1 and SNMPv2 based agents are widely implemented at the tested NRN edge devices: CISCO LS1010, CISCO LS100, FORE ASX200, UB GeoSwitch 155 and CISCO routers. An important number of ATM based standard MIBs, like the ATM MIB (IETF RFC 1695), PNNI-MIB, ATM Forum MIB, are widely supported and tested. Also, a large number of proprietary MIB extensions for ATM were tested. But it is evident that the proprietary MIBs are being replaced by standardized MIBs such as PNNI-MIB or ATM-RMON-MIB for statistics. But still there are too many proprietary MIBs which complicate the development of standard management applications for ATM networks. By a subset of the statistics offered by the largely implemented MIBs it was possible to estimate some Quality of Service, like the Cell Error Ratio (CER), on the VC level offered by the ATM Overlay network. The estimation showed that the CER is comparable with the default objective values for public services.

The OAM F4 and F5 Loopback flows (ITU-TS I.610) were tested at the CISCO ATM switches, as it had the proprietary solutions in MIBs. The Loopback cells within CISCO-OAM-MIB could serve for the management applications on connectivity check on VP and VC connections. Here, again, the problem of interoperability exists, as the MIB for OAM is not yet standardized.

Web based management is still very promising, but again we need standards. The Web page which was created for the ATM Overlay network status monitoring and the join test with JAMES on the Customer Management Interface (CMI) show all advantages of the Web based interfaces, but both were just prototypes.

The tests on remote PVC configuration using standard ATM MIB were successful what is important not only for management of LANs but also of ATM WANs.

### Participants

ACOnet (AT), ULB/VUB (BE), CERN (CH), SWITCH (CH), DFN (DE), UNINETT (NO), SURFnet (NL), RedIRIS (ES), GARR (IT), UKERNA (UK), RCCN (PT), NTU (GR) and CESNET (CZ).

### Dates et phases

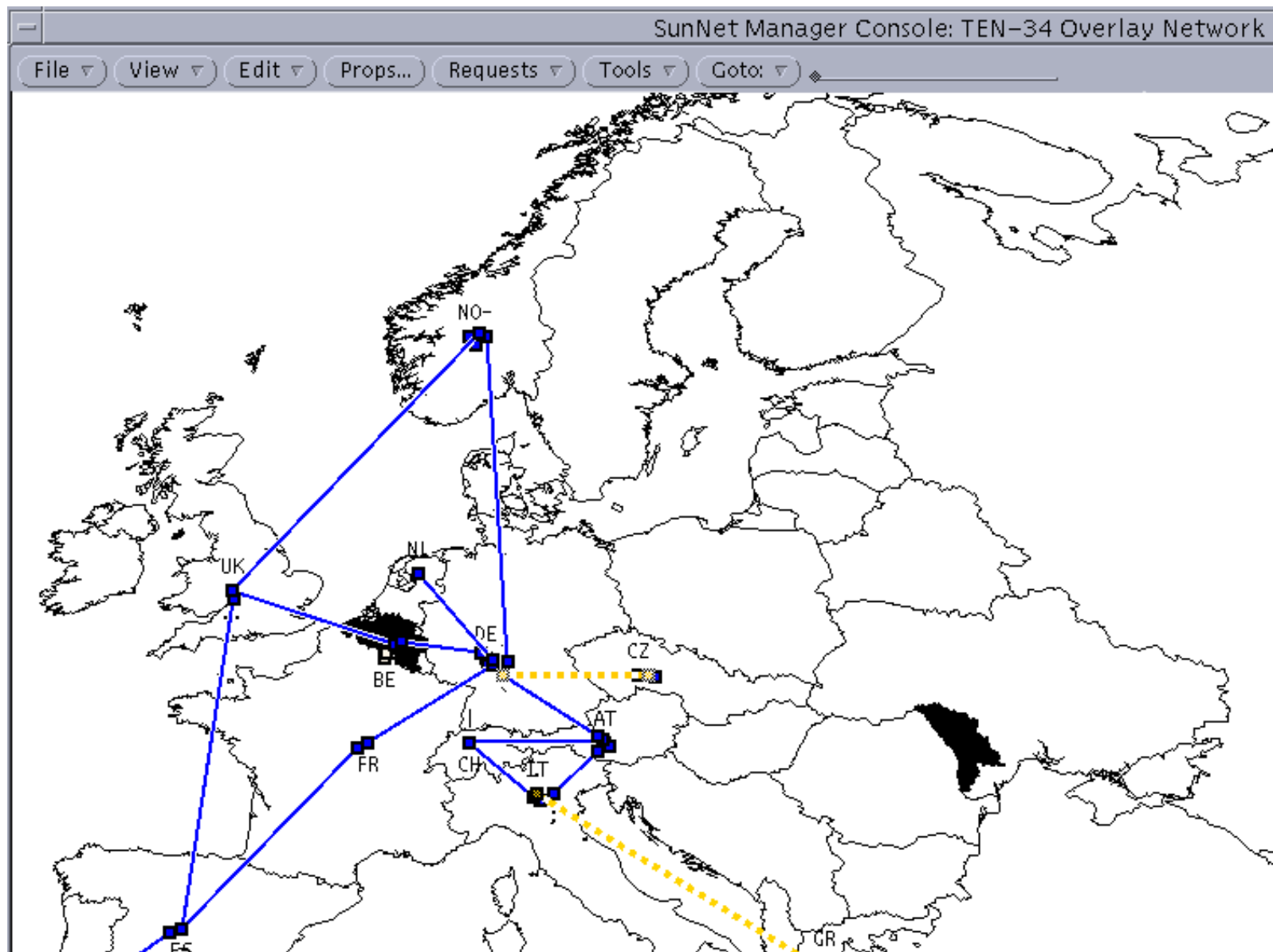
In general phases started as it was proposed in D14.1. The majority of the Network Management tests were performed continuously over the period of July '97-April '98. However, the tests on OAM flows, CMI and MIB based PVC configuration were realized in February and March '98.

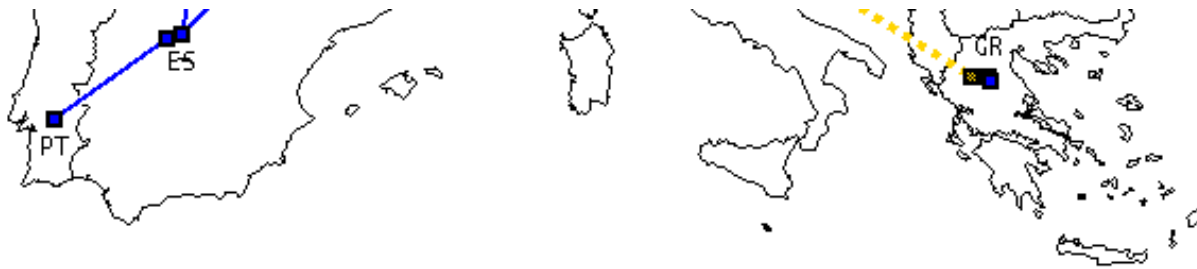
## Network infrastructure

The existing ATM Overlay network (User Information transport network) based on 2 Mbit/sec VP service from JAMES with NRN ATM edge devices was used for the tests. Two infrastructures were used for the management transport network: The Internet and the ATM Overlay network. A special configuration of 10 PVCs through the ATM Overlay network for the management tests has been realized.

NRN ATM equipment participated in the tests included:

- NORDUnet: Norway, Oslo, CISCO ATM switch LightStream1010 (LS1010) and CISCO router,
- GARR: Italy, Milan, FORE ATM switch ASX200 and CISCO ATM switch LightStream1010,
- ULB/STC: Belgium, Brussels, CISCO ATM switch LightStream100 and CISCO router 7010,
- ACOnet: Austria, Linz, CISCO ATM switch LS1010,
- SWITCH: Switzerland, Zurich, CISCO ATM switch LightStream 1010 and CISCO router,
- SURFnet: Netherlands, Twente, UB GeoSwitch 155,
- DFN: Germany, Stuttgart, CISCO ATM switch LightStream1010,
- UKERNA: United Kingdom, London, FORE ATM switch ASX200,
- RCCN: Portugal, Lisbon, FORE ATM switch ASX200 and CISCO ATM switch LightStream1010,
- RedIRIS: Spain, Madrid, CISCO ATM switch LightStream1010,
- RENATER: France, Paris, CISCO ATM switch LightStram1010.





*Fig.1: ATM Infrastructure used for Network Management tests*

## Local infrastructure

SunNet Manager-SunNet Domain Manager version 2.3 on Solaris 2.4 was used as a Management Platform for Network monitoring. It was connected both to the ATM Overlay Network and Internet network for the tests. For the Web page on the Overlay Network status monitoring the Web server with Windows NT at the ULB/VUB premises was used.

## Hardware/software

On the NRN edge devices: The production and early deployment releases of software which supported the latest standards were installed, like MIB II and SNMPv2 and AToM-MIB, ATM-RMON-MIB, PNNI-MIB.

The tests have been done on the CISCO LS1010, Version IOS (tm) from 11.1 up to LS1010 WA4-2 Software (LS1010-WP-M) Version 11.3 (0.8), SysObjectID=1.3.6.1.4.1.9.1.107.

CISCO routers were with versions of IOS 11.1 up to IOS (tm) RSP Software (RSP-JSV-M) Version 11.2 (11)P, Release (fc1), SysObjectID=1.3.6.1.4.1.9.1.48 were available.

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

## Results and findings

A specific management information flow (M1-M5), defined by the ATM Forum (1), includes a conceptual view and a MIB (Management Information Base) for the five different management interfaces. Private ATM network management is addressed through M1 combined with M2. M1 is concerned with management of end user equipment connecting to either private or public switches, and M2 with management of ATM switches and networks. M3 is the link between private and public networks to exchange fault, performance and configuration information. M4 pertains to management of public ATM switches and networks. M5 supports interactions or exchange of management information between any two public networks.

According to the M3 specification (2) "read only" management service (Class I of requirements) is mandatory if the service provider offers any management service.

Class I of requirements includes access and read of general, configuration and status information at the customer portion of the ATM public network.

Class II of requirements is optional for the service providers. It includes addition, modification or deletion of virtual connections and subscription information in a public network.

The JAMES didn't offer services from the Class 1 and 2, but our test scenario included the similar functions on M1 and M2 views, which if standardized could cover the M3 view. We used the SNMP based management as it represents the dominant protocol at the user side (management applications) as well as at the ATM equipment vendor side (agents). The following overview gives the SNMP applicable MIB extensions for the ATM available as standards during the phase 2:

- M1: AToM MIB (ATM MIB), LANE MIB, DXI MIB, Proprietary MIBs
- M2: AToM MIB, LANE MIB, ILMI MIB, CES MIB, PNNI MIB, Transmission MIBs, (RFC 1406, RFC 1407, RFC 1595) IMA MIB, ATM RMON MIB, Proprietary MIBs
- M3: M3 MIB, AToM MIB
- M4: ILMI MIB, LANE MIB, CES MIB, M4 MIB, AAL MIB, Transmission MIBs (RFC 1406, RFC 1407 and RFC 1595), IMA MIB, ATM RMON MIB.

In general, all these MIBs were of our interest and majority were available at the NRN edge devices.

However, some of them like CES (Circuit Emulation Service) and IMA (Inverse Multiplexer for ATM) were not yet implemented by the vendors as agents.

The test scenario (as defined in D14.1) consisted of four separated test groups:

1. Operation, Administrations and Maintenance Flows (F4/F5 OAM flows)
2. Web Based Management
3. SNMP Based Management tests
4. Xuser Based Management tests.

Further on, each of these groups of tests is described in more details.

### **Operation, Administrations and Maintenance Flows (F4/F5 OAM flows)**

OAM flows can be applied both at the physical and at the ATM layer, ITU-TS I.610 (7). The flows (F1, F2, F3) at the physical layer (F1, F2, F3) are dependent of the transmission system (SDH, PDH) and were not of our interest. At the ATM layer two flows: F4 and F5 are covering VP and VC level, respectively. Both flows are bidirectional and follow the same route as the user-data cells, thus constituting an in-band maintenance flow. Both ATM layer flows can either cover the entire virtual connection (End-to-End flow) or only parts of the virtual connection (Segment flow). Through the OAM flows, the following groups of functions could be realized:

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

Not all these functions are standardized and implemented. The loopback OAM flows are the first being standardized and implemented and therefore they were of our primarily interest. The loopback tests enable the verification of ATM Layer connectivity existence for a particular connection. For F4 flow VPI corresponds to tested VP and VCI is constant, always set to 4. For F5 flow VPI and VCI correspond to tested VC. The mechanism consists of sending out the loopback cells and activating timers. If the originator receives back the looped cells in the interval of 5 seconds it is assumed that the connectivity exists. In practice the OAM flows could be activated through a customer access UNI or through the TMN (Telecommunications Management Network).

For the OAM loopback flows the switches have to perform very simple checking: on loopback type (End-to-End/Segment), on indicator (forwarding flow/backwarding flow), on correlation tag (unique id. flow number) and on connection end-point location identifier. The last parameter is used in the loopback tests variant called Loopback Test Using Loopback Location Identifier as in the case of CISCO implementation. The Loopback Location Identifier is not standardized and at CISCO it is an ATM switch address prefix. The experience with OAM flows shows that the mechanism even in its early implementation phase is very promising manner to learn about ATM layer behavior throughout the large ATM networks.

In the Phase 1 the OAM tests were performed in the CISCO cloud with End to End Loopback tests activated from the management console system.

In the phase 2, the aim was to investigate the MIBs which support the managed objects for the F4/F5 flows and to test their applicability for the connectivity check. There exists no standardized MIBs for the F4/F5 flows. However, we tested the proprietary

MIBs on CISCO LS1010 which support the OAM flows through the CISCO-ATM-CONN-MIB and CISCO-OAM-MIB. The two types of functional tests concerning these two MIBs in the CISCO were done in March '98. The tests were "functional" i.e. the aim was to investigate the possibility to create applications on OAM Loopback managed objects which will enable an on-line connectivity check.

The tests have been done on the CISCO LS1010 , Version IOS (tm) LS1010 WA4-2 Software (LS1010-WP-M) Version 11.3 (0.8), SysObjectID=1.3.6.1.4.1.9.1.107. The switch was located at SWITCH, Zurich, while the SNM platform from Brussels used the SNMP "set" request. The "write community string" was enabled for the SNM platform at the switch. For the transport network we used Internet, even, there would be possible to use the ATM PVC if we changed that the PVC ended at the ATM switch, but not at the router (as was realized).

#### ATM Loopback activation with CISCO-ATM-CONN-MIB

The test covered the activation of loopbacks and generation of OAM connectivity cell flows at VP and VC using CISCO-ATM-CONN-MIB.

The activation of OAM F4 flow was done at the port 5, VPI 6, choosing the interval 5 (frequency of OAM loopback generated cells). Set "enable end2end OAM loop" and "enable segment OAM loop" were tested. The activation of OAM F5 flow was done at the port 2, VPI 0, VCI 164 i.e. on the PVC used for the Overlay network status monitoring, see Fig. 2.

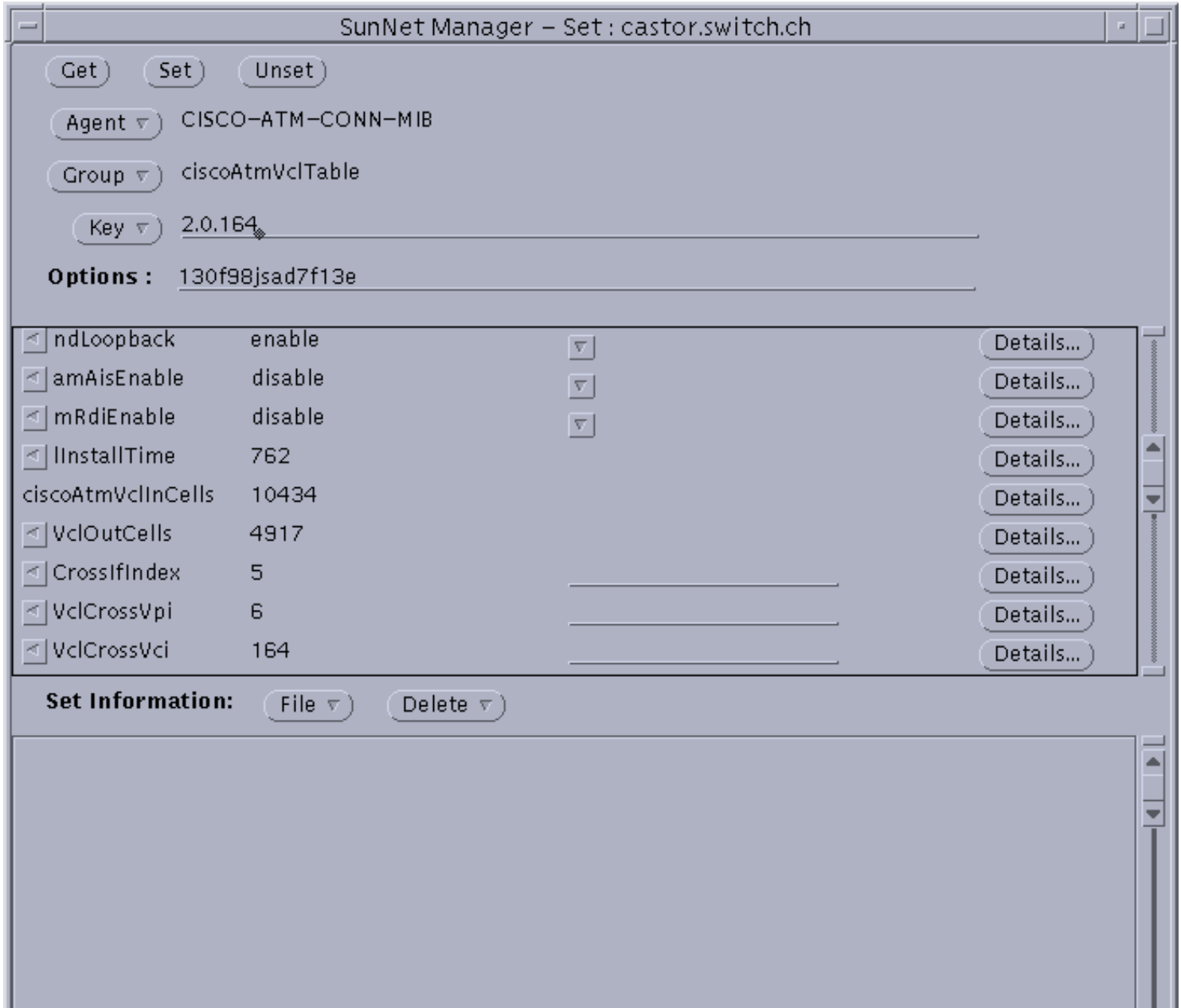




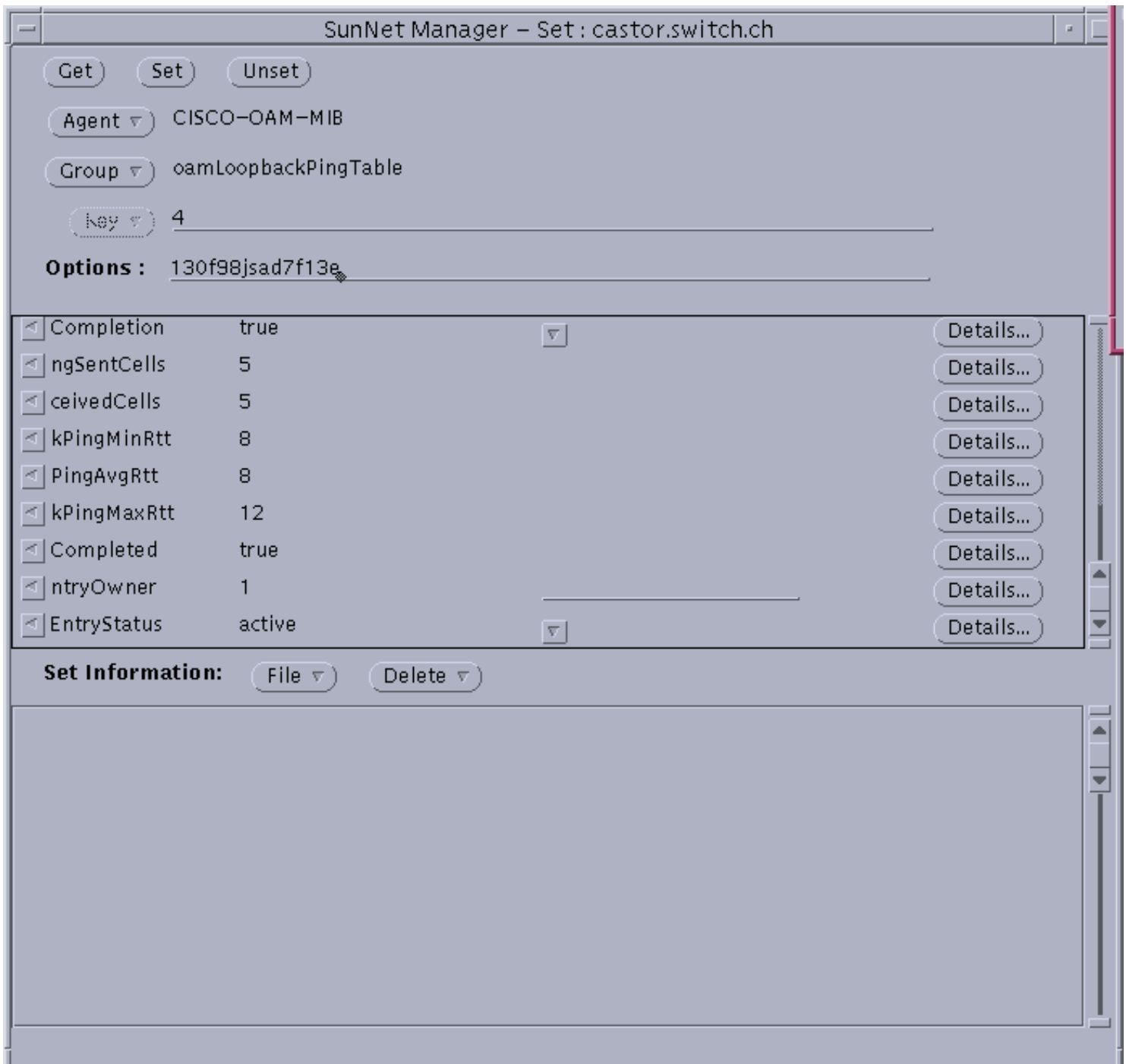


Fig. 2. Remote SNMP "set" of F5 OAM Loopback flow (CISCO-ATM-CONN-MIB)

### ATM Loopback tests with CISCO-OAM-MIB

The test covers the SNMP based creation of test table entries and control of OAM test results using CISCO-OAM-MIB.

This, second test, used more complex procedures as it was necessary first: to activate the oamLoopbackPingTable entry, second: to specify the details about OAM flows for a specific VP or VC and third: to browse the results. In CISCO realization after the results are obtained table is available 5 minutes and after that deleted. As illustration of the the procedures see Fig. 3.






Fig. 3. Remote oamLoopbackPing Table result browsing (CISCO-OAM-MIB)

These tests showed that this type of test tables (similar test tables are in the standardization procedure by IETF (15), could be used as a principal method for the connectivity check in ATM networks and for the outage statistics.

## Web Based Management

The Web based management functions were tested in two scenarios. One was the ATM Overlay Network status monitoring and the other was the CMI (Customer Management Interface) joint test with JAMES. The later is explained under the title *Xuser Based Management tests*.

For the monitoring we used two systems in parallel:

A Web based system for the public on-line status monitoring and, separately, the SunNet Manager platform both for the status monitoring and statistics collection. Beside this the Web based browsers for SNMP requests and CMIP were available on the Web page.

### Web page for Overlay Network status monitoring

The Web based page was realized in January '98 and was used for monitoring of 13 VPs in the Overlay network through 10 PVCs inside them. The aim was to cover the 15 following VPs: BE-DE, BE-UK, DE-NO, DE-NL, DE-AT, AT-CH, AT-IT, CH-IT, IT-GR, NO-UK, UK-ES, ES-PT, DE-FR, FR-ES, DE-CZ. In the phase of design we calculated that for those VPs in the configuration (with CZ and GR in) we will need 53 PVCs and 7 subnetwork address spaces to cover perfectly the topology. As our aim was not to create a perfect application we decided to create an optimal number of PVCs which will cover all VPs having in mind that the VPs normally have to be "in service" and that additional tests could be done to analyze specific service outages in detail. From the central management platform in BE, we used two VPs: BE-DE and BE-UK which were back-ups of each other. The other PVCs which were monitored in the period January-March '98, had the following characteristics (The VPI numbers are not presented in the list as they were only of local importance):

PVC001:\*\*\*\*\*BE\_UK (VCI=132, IP address=192.168.0.132)

PVC002:\*\*\*\*\*BE\_DE (VCI=161, IP address=192.168.0.161)

PVC003:\*\*\*\*\*BE\_DE\_FR\_ES\_PT (VCI=171, IP address=192.168.0.171)

PVC004:\*\*\*\*\*BE\_DE\_AT\_IT\_CH (VCI=164, IP address=192.168.0.164)

PVC005:\*\*\*\*\*BE\_DE\_AT\_CH\_IT (VCI=165, IP address=192.168.0.165)

PVC006:\*\*\*\*\*BE\_DE\_NO\_UK\_ES (VCI=169, IP address=192.168.0.169)

PVC007:\*\*\*\*\*BE\_DE\_AT (VCI=168, IP address=192.168.0.168)

PVC008:\*\*\*\*\*BE\_DE\_FR (VCI=175, IP address=192.168.0.175)

PVC009:\*\*\*\*\*BE\_DE\_NL (VCI=173, changed in DE to 131, IP address=192.168.0.173)

PVC010:\*\*\*\*\*BE\_DE\_NO (VCI=174, IP address=192.168.0.174).

The range of attributed VCI was between 132-175, even it could start with 32. But some NRNs had already in use the range up to

100. The range was less than 255 as we wanted to make the correspondence between the VPI and the end-system IP address from the private subnet domain (192.168.0.0) for the simplicity reason.

The PVCs end-points were configured to finish in switch, router or workstation, what was the free choice for the participants.

Only in one case (NL), the VCI from one end-point was changed before the other end-point. The VCI 173 configured between BE and DE, is changed to be 171 between DE and NL, because of constraints to use this VPI number at the UB GeoSwitch 155.

Configuration has been done in three steps. First, the configuration of 2 "one-hop" PVCs has been done (PVCs BE\_DE, BE\_UK) and after the verification of their functionality the second step has been done. The second step included new 2 "two-hop" PVCs (BE\_DE\_NO and BE\_UK\_ES). Then, all others were configured and tested in parallel. The questions and answers of PVC configuration at the ATM switches and routers were sent to the e-mail list, so after 4 first PVCs no questions nor problems concerning configuration were experienced. One question arised in the case were an ATM edge device was accessed both by the Internet and the ATM PVC: how the PING echo will be returned back to SNM platform, as it had the same operational IP address in two configuration tables (in the routing table for the Internet and in the ATM PVC mapping table). Finally, we agreed that it wouldn't disturb the monitoring function as long as we know what is happening.

The status Monitoring Web page used a freeware shell script tool ("Big Brother" software) which consists of few parts:

- Central monitoring station (Display Server). This station accepts incoming reports and prepares them for display. The display matrix shows a status of green (ok), yellow (warning), red (severe), and blue (no contact) for each system/area combination. Furthermore, the entire screen changes color to reflect the most serious condition on the network. In order of increasing severity these conditions are: green, yellow, blue, red.
- Network monitor. The network monitor periodically contacts every ATM switch in the PVC list file via IP PING through the ATM PVC. Results are then sent to the system designated as the Display Server. The polling period was 5 minutes.
- Pager Programs. The client which sends single lines of information to the designated server and executes a script (page) which forwards this information using Kermit via modem to the designated pager. This option was not activated in the test but was present as indication of useful possibility in the monitoring of this type.

In general, this Web page, although very elementary, helped all participants to check the ATM network on-line. Psychologically, it motivated the participants to announce all planned outages of their ATM edge systems and to explain the unexpected ones. That was not the case in the phase 1, when we were without the publicly available network control system. The Fig. 4 shows the Web page.

**Legend**

- ATM Virtual Path/PVC OK
- Attention
- Trouble
- No report

**Updated**  
Sat Jan 31 11:14:05 MET  
1998

**CONN**

as12	●
PVC001:*****BE_UK	●
PVC002:*****BE_DE	●
PVC003:*****BE_DE_FR_ES_PT	●
PVC004:*****BE_DE_AT_IT_CH	●
PVC005:*****BE_DE_AT_CH_IT	●
PVC006:*****BE_DE_NO_UK_ES	●
PVC007:*****BE_DE_AT	●
PVC008:*****BE_DE_FR	●
PVC009:*****BE_DE_NL	●
PVC010:*****BE_DE_NO	●

Buttons: Help, Info, Page, View

Fig.4. Overlay Network status Web page with no alarms

### SNMP Based tests with SNM platform

As the Web page for the Overlay network status monitoring is memoryless (i.e. it collects no logs) we used the more sophisticated system for the statistics purposes. The statistics collection, based at 15 minute polling period, for the connectivity check was realized by the SNM platform. The platform was available for the interested participants from other NRN networks through remote X window sessions as the SNM allows simultaneous usage of different management views.

Here, four types of tests were done:

1. SNM View for Overlay network status monitoring
2. MIB access and read
3. Cell Error Ratio (CER) estimation
4. PVC configuration using MIBs.

Further on, these four tests are explained.

### SNM View for Overlay network status monitoring

In the period January '98 - April '98 the SNM platform collected the IP PING Echo packets obtained as a result of the PINGs generated each 15 minutes through the 10 PVCs. The requests contained 5 PING packets of 64 octets.

The collected statistics on connectivity for PVC BE\_UK and BE\_DE\_NO is presented in Fig. 5 and Fig. 6.

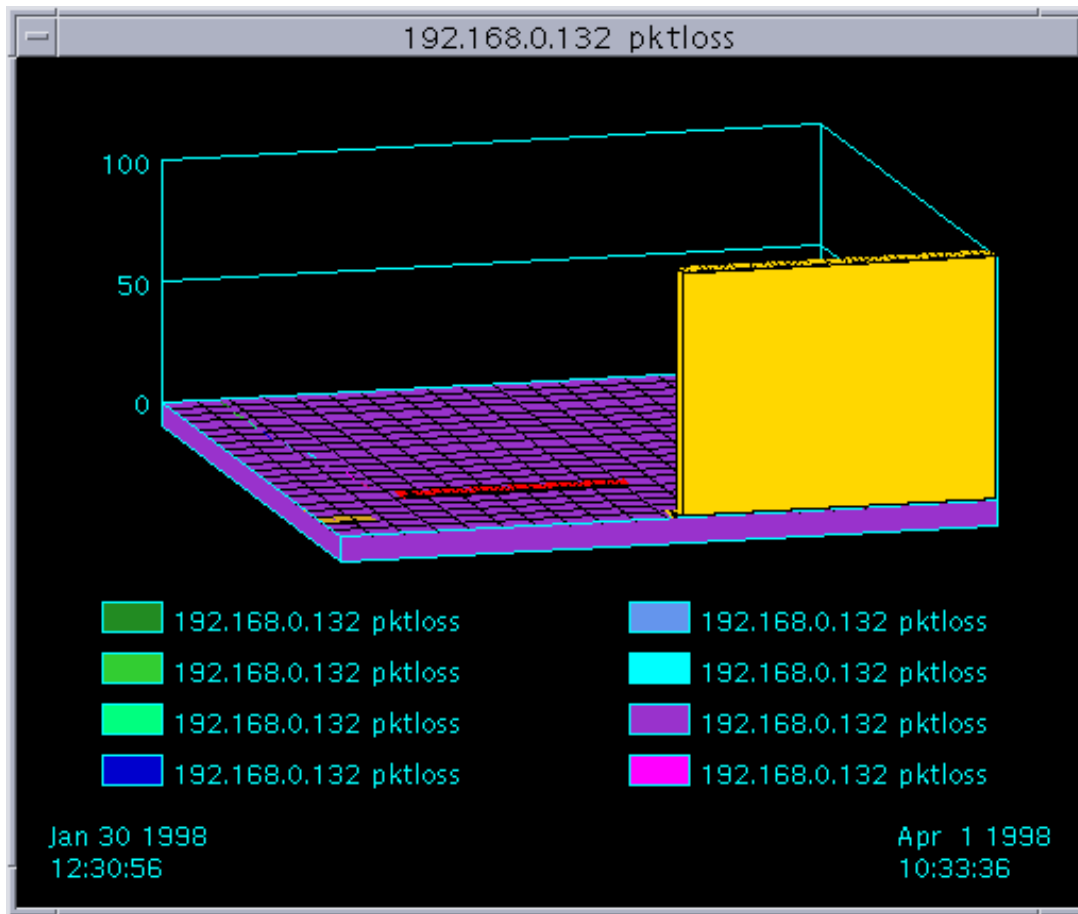


Fig. 5. Connectivity statistics for PVC 1 (BE\_UK)

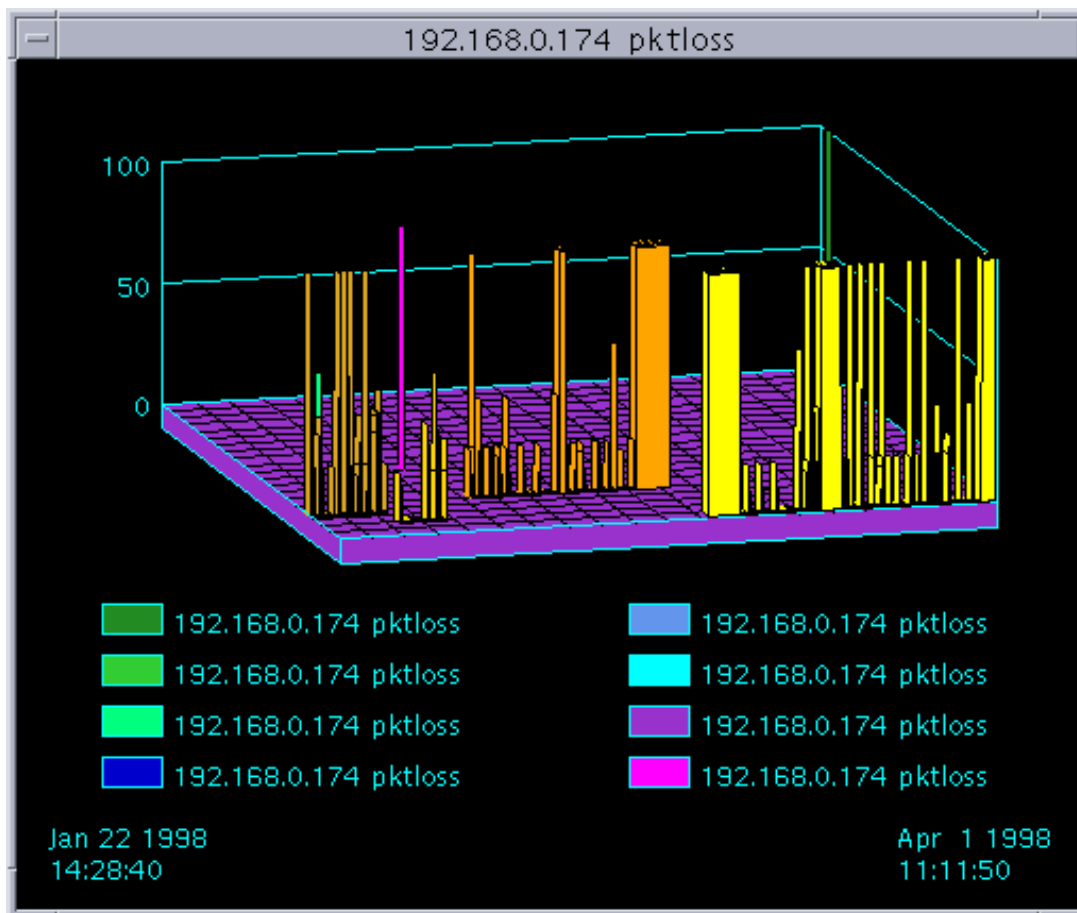


Fig. 6. Connectivity statistics for PVC 10 (BE\_DE\_NO)

In the very beginning we had a problem at the PVC BE\_UK where the loss of IP packets were about 90 %. With a new version of software for the FORE ASX200, in January 1998, the problem was solved, and this PVC had an excellent quality in the period it was active. A non-active period for this PVC is due to the PNNI tests which requested this VP to exclude from the NRN switch. The main reason for the periods of nonactivity at the other PVCs were the end-system problems, but some of them were suspicious to be generated by the VP status itself, see Fig. 6. That was not possible to verify without analyzers.

#### MIB access and read

The management view for this test was created already in 1996 and was active all the time till the end of March '98. It was periodically updated by the new OID and schema files as new software releases for the switches and for the routers included agents with new MIBs.

For the majority of "MIB access" tests we used the CISCO cloud. In FORE cloud we had 2 FORE ASX200 (in PT and UK) and 1 UB GeoSwitch 155 in NL. In CISCO cloud we tested software versions IOS from 11.1 up to 11.3. The latest version of CISCO LS1010 was: IOS (tm) LS1010 WA4-2 Software (LS1010-WP-M) Version 11.3 (0.8), SysObjectID=1.3.6.1.4.1.9.1.107.

The latest version of CISCO router was: IOS (tm) RSP Software (RSP-JSV-M) Version 11.2 (11)P, Release (fc1), SysObjectID=1.3.6.1.4.1.9.1.48.

For the statistics collection we used SNMPv1 "read community string" management functions both for SNMPv1 and SNMPv2 agents. These tests included SNMPv1 read access and statistics collection for different MIB extensions for ATM. The MIBs of interest were the traffic statistics per interface and connection and AAL5 errors. This statistics was used for the estimation of certain parameters of Quality of Service offered by the JAMES network.

The SNM (SunNet Domain Version 2.3 on Solaris 2.4) was used for the access to the MIBs in eleven locations as follows:

- NORDUnet: oslo-atm.uninett.no (128.39.2.19) read community string: public; access over Internet,
- INFN: miasx200.mi.infn.it (192.84.138.200) read community string: ten-34, LS1010.mi.infn.it (192.84.138.11) read community string: ten-34; access over Internet,
- ULB/STC: rtr02.iihe.ac.be (193.190.246.65) read community string: public; write community string; access over Internet,
- ACOnet: jkuatmt1.edvz.uni-linz.ac.be (140.78.2.102) read community string: TEN- 34; access over Internet,
- SWITCH: popocatepetl.switch.ch (130.59.16.213) read community string: ten-34, write community string; castor.switch.ch (130.59.16.6) read comm. string: ten-34; access over Internet,
- SURFnet: atms2.cs.utwente.nl (130.89.10.230) read community string: tf-ten; access over Internet,
- DFN: ksatm3.rus.uni-stuttgart.de (193.196.152.2) read community string: tf-ten-nm; access over Internet,
- UKERNA: lemon.ukerna.ac.uk (193.246.0.226) read community string: public; access over ATM PVC;
- RedIRIS: atm-sw.rediris.es (130.206.1.43) read community string: public; access over ATM PVC;
- RCCN: marte.rccn.net (193.136.7.34), 193.136.7.71, read community string: public, access over Internet;
- RENATER: 192.93.173.243, read community string: JAMESMNGT, access over Internet.

Different standard "community based" protection levels were used for SNMP read access: public and group community strings. In the case of UKERNA and RedIRIS, the SNMP access was realized only through the tunneled PVC over the ATM Overlay network because of firewall, which doesn't allow the UDP packets to pass through.

This SNM view, again, was possible to test remotely through the X window sessions.

The following experience concerning MIBs at CISCO LS1010 was gained:

The *CISCO LS1010* implements both the standardized ATM-specific MIBs: RFC 1695, ATM PNNI

The following support was available in Release **11.1 of the LS1010** software.

- **Standardized ATM-specific MIBs:** The LS1010 implements both of the presently standardized ATM-specific MIBs: RFC 1695, ATM PNNI MIB and the ATM LANE Client MIB.

**ATM-MIB:** An Internet Standard MIB (AToM MIB, RFC1695) for basic management of ATM interfaces and VCCs in ATM end-systems and ATM switches,

**PNNI-MIB:** Only pnniRouteAddrTable,

**LAN-EMULATION-CLIENT-MIB:** The ATM Forum's LAN Emulation Client MIB.

- **Number of LS1010-specific MIBs:**

**CISCO-ATM-ADDR-MIB:** The MIB contains a list of the valid calling party

addresses for a UNI on a per interface basis,

**CISCO-ATM-CONN-MIB:** The MIB module for an extension to RFC1695 VPL,VCL tables and for SVP, SVC ATM address tables,

**CISCO-ATM-IF-MIB:** The MIB module for an extension to RFC1695 ATM interface tab,

**CISCO-ATM-IF-PHYS-MIB:** Subset of SONET and DS3 MIB for LS1010,

**CISCO-ATM-RM-MIB:** The MIB complements standard ATM MIBs for Cisco ATM switch Resource Management,

**CISCO-ATM-SERVICE-REGISTRY-MIB:** A MIB module to allow an NMS to monitor and configure the information which an

ATM switch makes available via the ILMI's Service Registry Table,

**CISCO-ATM-SWITCH-ADDR-MIB**: ATM Switch address MIB,

**CISCO-ATM-TRAFFIC-MIB**: The MIB for an extension to traffic OIDs and variables defined in RFC1695,

**CISCO-RHINO-MIB**: The Chassis MIB for LS1010 ATM switch,

The following additional support was available from **Release 11.2. for LS1010**:

- **LS1010-specific MIBs:**

**ACCOUNTING-CONTROL-MIB**: The ATM Accounting Control MIB, from draft-ietf-atommib-acct-04.txt,

**ATM-ACCOUNTING-INFORMATION-MIB**: ATM Accounting Information MIB, from draft-ietf-atommib-atmacct-01.txt,

**ATM-RMON-MIB**: ATM Remote Monitoring MIB, from an ATM Forum draft - AF-NM-TEST-0080.000,

**CISCO-ATM-ACCESS-LIST-MIB**: A MIB for configuration and control of access control filters in an ATM switch,

**CISCO-OAM-MIB**: The MIB invoking OAM loopback Ping on ATM connections,

**CISCO-PNNI-MIB**: Cisco specific extensions to the ATM Forum PNNI MIB,

**PNNI-MIB**: The MIB for managing ATM Forum's PNNI routing protocol with additional support for the PNNI-MIB:

pnniBaseGroup, pnniNodeTable, pnniNodeTimerTable,

pnniScopeMappingTable, pnniSummaryTable, pnniIfTable, pnniLinkTable,

pnniNbrPeerTable, pnniNbrPeerPortTable, pnniRouteAddrTable.

For the Cisco routers a dozens MIBs which are not directly related to ATM interface existed but they are not all relevant here. ATM related MIBs which were of interest are cited below.

**CISCO Routers SNMP MIBs:**

- **SNMP version 2 MIBs in IOS 11.0:**

**CISCO-QUEUE-MIB**: The Cisco Queue MIB. This MIB displays the queue statistics reported by "show queueing" and "show interface",

**IPMROUTE-MIB**: The IPMROUTE MIB: from draft-ietf-idmr-multicast-routmib-00.txt,

**PIM-MIB**: The PIM router MIB: from draft-ietf-idmr-pim-mib-00.txt,

**PNNI-MIB**: The MIB for managing ATM Forum's PNNI routing protocol.

- **SNMP version 1 MIBs added in IOS 11.1:**

**RMON-MIB**



- **SNMP version 2 MIBs added in IOS 11.1:**

**CISCO-LECS-MIB:** The Cisco MIB for creating, configuring and monitoring the LAN Emulation Configuration Server as well as entering/modifying data within the the LECS database,

**CISCO-LES-MIB:** The Cisco MIB for creating/monitoring the LAN Emulation Server. This MIB also allows the monitoring of the LANE clients as perceived from the server,

**CISCO-MEMORY-POOL:** (added in 11.1(2)): The Cisco MIB for monitoring memory pools,

**LAN-EMULATION-CLIENT-MIB:** The ATM Forum's LAN Emulation Client MIB.

- **SNMP version 2 MIBs added in IOS 11.2:**

**ATM-MIB:** IETF AToM MIB

Only groups atmInterface ConfTable, TrafficDescrParamTable and atmVclTable were tested as other groups were not supported at our versions of routers.

In FORE ASX200 cloud the following MIBs were tested:

- **FORE ASX200** with FT5.1 with following SNMP MIBs:

**FORE-SWITCH.MIB:** Smart Permanent Virtual Circuits statistics and manipulation /PNNI SPVC,

**RFC1573.MIB:** Interface Group MIB - II, IANAifType-MIB DEFINITIONS,

**RFC1213.MIB:** mib-2,

**PNNI.MIB:** From file : "af-pnni-0081.000.mib",

**ATM-FORUM-ADD-REG-MIB:** Address registration MIB.

At UB GeoSwitch 155 we tested:

**ATM-FORUM-ADD-REG-MIB:** Address registration MIB,

**ATM-MIB:** IETF AToM MIB.

The tests done in March '98 covered also the activation of ATM-RMON-MIB at CISCO LS1010 (CH). portSelTable and portSelGrpTable with 2 entries were created but it was noticed that all groups in the MIB could not be activated. The reason could be that this MIB needs more memory at the switch.

#### **Cell Error Ratio (CER) estimation**

The monthly statistics on Interface traffic and AAL5 errors were put at the Web page for the majority of NRN ATM switches. First the statistics for the AConet, DFN, GARR, NORDUnet, SURFnet and SWITCH was collected and later for the RENATER and RCCN was added as they were connected to the Overlay network. UKERNA and RedIRIS were included the latest as it was necessary to use the ATM PVCs for SNMP requests because of the firewalls in these NRN (firewalls don't allow UDP packets for SNMP).

The ATM Overlay network, designed over the JAMES ATM infrastructure, had a class U of service specified at the NRN ATM

edge devices. "U" means "unspecified" or "unbounded". For the U class of service ITU-T establishes no objective for CER (Cell Error Ratio) because OAM performance monitoring capabilities are ineffective in class U. However, for the classes "1" (stringent class), "2" (tolerant class) and "3" (bi-level class), the upper bound on the cell error probability is set to  $4 \times 10^{-6}$ . This Cell Error Ratio represents a default objective for the public B-ISDN on a 27 500 km hypothetical reference connections, according the ITU-T Recommendation I.356 (9). In the same Recommendation it is said that it is expected that the national and international portions will achieve their allocations for CER even for QoS class U.

Cell Error Ratio (CER) is the ratio of total errored cells to the total of successfully transferred cells, plus tagged cells, plus errored cells in a population of interest. Successfully transferred cells, tagged cells, and errored cells contained in severely errored cell blocks are excluded from the calculation of cell error ratio.

Cell Error Ratio can be measured out-of-service by transferring a known data stream into the network at the source measurement point (MP) and comparing the received data stream with the known data stream at the destination MP. It has been suggested that a BIP 16 (Bit Interleaved Parity) indicator at the Physical layer using small blocks (less than 200 cells) could be used as CER is primarily governed by transmission performance.

Estimation of Cell Error Ratio by in-service measurement is desirable but difficult. Having in mind the absence of all these complicated mechanisms and the absence of error control covering the cell information field, we assume that the AAL5 error counters could help in Cell Error Ratio estimation at the end-to-end ATM connections.

In our calculations, the AAL5 errors are taken from the ATM-MIB (AToMIB) AAL5 Connection Performance Group which includes: AAL5 CPCS (Common Part Convergence Sublayer) PDU error counters of CRC-32 errors, SAR (Segmentation and Reassembly) time-outs and oversized SDU (Service Data Units). For the CER only CRC errors were of interest.

The statistics in the Tables 1-6 show the average number of ATM virtual connections, traffic and the AAL5 errors at interfaces/ports of interest in some NRN networks. We estimate that one error at the AAL5 layer corresponds to one errored cell (if there are burst errors they are not calculated anyway in the Cell Error Ratio). These ATM switches participated in the tests with mutual exchange of traffic over the ATM network with distances of a few thousand kilometers passing international borders. End-to-end estimation of CER objectives for the TEN-34 Overlay network (according to allocation rules of Recommendation G.826) showed that the complexity of this Overlay network is rather close to the reference hypothetical network from ITU-T Recommendation I.356 (9).

The Statistical Cell Error Ratios from the Tables 1, 3 and 4 estimate low CER (in AConet, GARR, NORDUnet), even better than the default objective for public services. In two cases (DFN and SURFnet), Table 2 and Table 5, the statistical CER is higher, but the total number of received cells in these end-points is relatively small. At the other hand, the Table 6 (SWITCH) has no errors, but the number of received cells is also relatively small. In these last three cases we could pose the question of the statistical performance commitment for connections with small traffic during their lifetime.

The CER per VC is calculated as ratio of the AAL5 CRC-32 errors to the -Traffic- Received Cells related to the port to Average No. of VC per that port.

**Table 1. Statistics at AConet-port 1**

	Average No. of VCs	-Traffic- Received Cells in $10^6$	-Traffic- Send Cells in $10^6$	AAL5 errors	Cell Error Ratio per VC
April '97 6	47	380	720	53	$3 \times 10^{-6}$

May '97	35	140	150	415	$8 \times 10^{-8}$
June '97	38	1 530	10 500	647	$1 \times 10^{-8}$
July/Aug '97	17	610	580	830	$8 \times 10^{-8}$

---

**Table 2. Statistics at DFN-port 14**

---

	Average No. of VCs	-Traffic- Received Cells in $10^6$	-Traffic- Send Cells in $10^6$	AAL5 errors	Cell Error Ratio per VC
April '97	8	250	1320	0	0
May '97	7	2	2	4602	$9 \times 10^{-5}$
June '97 0	25	20	90	0	
July/Aug '97	29	30	10	0	0

---

**Table 3. Statistics at GARR-port 5**

---

	Average No. of VCs	-Traffic- Received Cells in $10^6$	-Traffic- Send Cells in $10^6$	AAL5 errors	Cell Error Ratio per VC
April '97	8	720	450	0	0
May '97	8	110	450	16	$7 \times 10^{-8}$
June '97	19	4500	4500	0	0
July/Aug '97	18	18000	9000	0	0

**Table 4. Statistics at NORDUNet-port 5**

	Average No. of VCs	-Traffic- Received Cells in 10 <sup>6</sup>	-Traffic- Send Cells in 10 <sup>6</sup>	AAL5 errors	Cell Error Ratio per VC
April '97	15	13500	15000	62	3 X 10 <sup>-10</sup>
May '97	18	7500	9000	227	8 X 10 <sup>-9</sup>
June '97	5	21000	24000	0	0
July/Aug '97	5	495400	48000	0	0

**Table 5. Statistics at SURFNet-port 1**

	Average No. of VCs	-Traffic- Received Cells in 10 <sup>6</sup>	-Traffic- Send Cells in 10 <sup>6</sup>	AAL5 errors	Cell Error Ratio per VC
April '97	3	1	1	0	0
May '97	3	1	1	22	2 X 10 <sup>-5</sup>
June '97	2	1	1	8	4 X 10 <sup>-6</sup>
July/Aug '97	2	4	2	665	4 X 10 <sup>-4</sup>

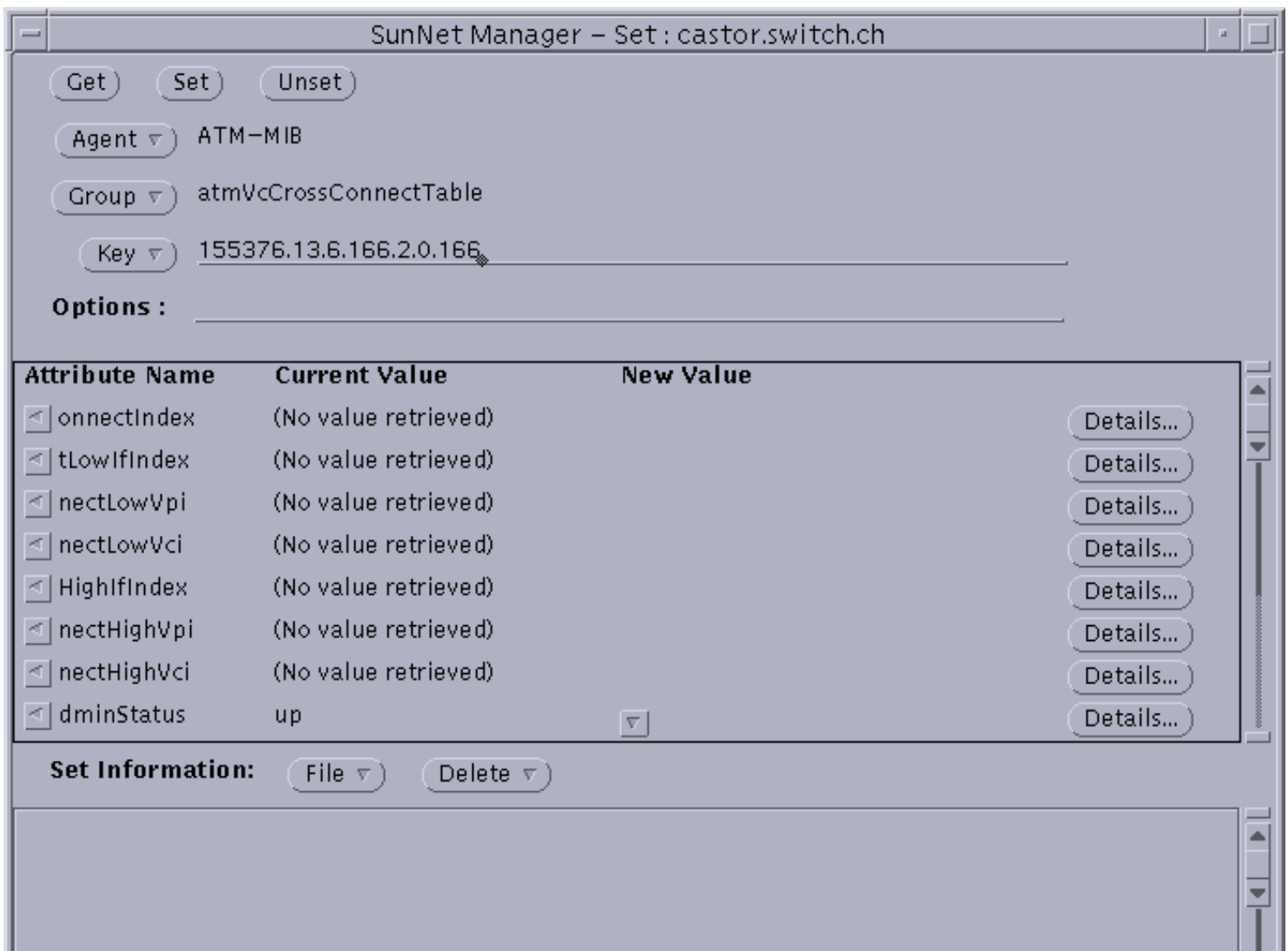
**Table 6. Statistics at SWITCH-port 8**

	Average No. of VCs	-Traffic- Received Cells in	-Traffic- Send Cells in	AAL5 errors	Cell Error Ratio per
--	--------------------------	-----------------------------------	-------------------------------	----------------	-------------------------

		10 <sup>6</sup>	10 <sup>6</sup>		VC
April '97	4	1	60	0	0
May '97	4	5	8	0	0
June '97	2	10	50	0	0
July/Aug '97	2	1	1	0	0

### PVC configuration using MIBs

These tests were done during March '98. For the PVC configuration we used the ATM-MIB (AToMIB) at the CISCO LS1010 switch (CH). Using the SNMP "set" from the Management platform in BE the PVC with VCI 166 at VP 6 (interface 13) passing to VCI 166 at VP 0 (interface 2) was created. For this test we needed three configuration steps. First, set the attribute "create and go" (or with "create and wait") in the vclRowStatus in the atmVclTable for two new entries ( 13.6.166 and for 2.0.166). Second, read the first free VC crossconnect index from the atmMibObjects group. Third, set the attribute "create and wait" in the atmVcCrossConnectTable as an entry with the parameters which connects the two PVC-ends and the VC crossconnect index. The two entries in the atmVclTable in ATM MIB have been then activated. The third step is presented in Fig. 7.





*Fig. 7. PVC configuration in the group atmVcCrossConnectTable from ATM-MIB*

## **Xuser based Management**

The Customer Management Interface (CMI) is a web-based User interface developed by WP4.2 for the JAMES project. It provides a direct interface between the customer and operator, and thus may reduce some currently employed, costly and wasteful fax procedures.

The CMI was initially proposed to be CMIP based and later changed to Web based.

It represents the interface between the Customer Management Domain and the Service Provider Management Domain enabling on-line provision of JAMES Technical Framework Document (TFD). Between the two Service Provider Domains, Xcoop interface, which was not the subject of the tests, is applied.

The TF-TEN common tests with JAMES on CMI were realized in two groups:

- Evaluation of CMI concept
- Simulation of real user scenario.

Further on, these tests and test results are explained in more details.

### **Evaluation of CMI concept**

The tests for evaluation were realized at the Web server in December 1997 and was initiated by JAMES (Reinhard Zagolla, DT). The following members of TF-TEN participated in evaluation of the CMI: CERN, RENATER, UKERNA, INFN-CNAF, SURFNET, NTU Athens, CESNET, VUB/ULB and DANTE.

All NRN participants expressed the positive opinion about the CMI. They find that the Berkom demo with the Web based interface is an excellent basis for dynamic bandwidth/connection configuration and for the status monitoring of the ATM networks.

As a conclusion, the TF-TEN proposed to test the CMI together with JAMES in the domain of Customer Control Panel for the needs of the TF-TEN Overlay network. This meant that we would like to use only one CMI Web page for it i.e. to have a centralized collection of requests if possible. This resulted in the second group of the tests which are described further on.

### **Simulation of real user scenario**

The test had two trials. The first trial was initialized by JAMES (Paul Richmond, BT) and it was planned to test the functions of ADDING, MODIFYING, DELETING and VIEWING in the domain of CMI user and services during the period January 1, 1998 and February 28, 1998. After the initial tests at the Web server, the JAMES side realized problems to control the Web server as the Service Provider and the test was postponed.

The second trial was performed between February 25 and March 4, 1998 and was realized as planned. The tests were performed using Web server.

The aim of the tests was to evaluate the functioning of the CMI in a simulated 'real' situation. In order to do this it was essential that both parties involved were independent of the development work group. This was to accurately test how new users would respond to such an interface and how easy it was to use first hand, for the first time, in a real situation. Paul Richmond played the part of the Lead Operator (CSR) and Zlatica (Zlatica Cekro-BE) was the JAMES user.

From the first unsuccessful trial, it was evident that the real life situation we hoped to create was not going to materialize. We concluded that our predetermined entry position for the second test trial was as follows:

1. Zlatica (user) has requested a JAMES connection and has completed a JUD (JAMES User Document) with the User Rep.
2. As the Lead CSR, Paul was passed the JUD from the User Rep and set up the TFD (Technical Framework Document).
3. As a result of this, the Users were contacted by the CMI to say the TFD had been set-up.

Further on, the complete Scenario prepared by JAMES (Paul Richmond, BT) is presented. After it, the description of test realization is presented in log form.

### **Adding - 25 Feb 98**

*Add new fields/data to the TFD.*

4.1 Zlatica to add user: R.Zagolla zagolla@telekom.dbp.de DT Germany

4.2 Paul to receive TFD addition notice successfully and notify Zlatica of success.

### **Deleting - 25 Feb 98**

*Delete existing fields/data from the TFD.*

5.1 Zlatica to delete user B, J.Aronsheim

5.2 Paul to receive TFD deletion notice successfully and notify Zlatica of success.

### **Modifying - 27 Feb 98**

*Change some of the existing details on the TFD.*

6.1 Zlatica to change cell rate to 7076 for CBR connection ID 2. (B-A)

6.2 - change the test date to 31/03/98, for connection ID 4 (VBR)

6.3 Paul to receive alteration requests successfully and report back to Zlatica.

### **7) Multiple changes - 2 March 98**

*This will be initiated by Paul since the previous test were from Zlatica, the user.*

7.1 Paul to request a whole new series of changes - without previously telling Zlatica.

7.2 Zlatica to report back on what modifications were made as a result of the e-mail notifications she receives. include contact user section.

## 8) Viewing - 3 March 98

This is a fundamental area of consideration and throughout the testing, check the usability of the pages - how can they be made more efficient and easier to use. Also comment on any 'viewing' ideas or events that happened throughout the project.

### Test Logs 24th February 1998

JAMES:

Testing began.

In order that the CMI could e-mail Zlatica as a user and Paul as a CSR, the TFD had to be modified first to include our details. If the TFD could have been set-up by the Paul (CSR) this would have already been done.

- users D and E were deleted, the text turned grey.
- two new users, F and G were created for Paul and Zlatica respectively, with the correct details in place. This text also turned grey.

Test plan was agreed.

**C**ustomer  
**C**ontrol  
**P**anel

**JAMES** *Joint ATM Experiment on European Services*

**User Information**

**JAMES Network View**

**Project Details**

Project Name	TEN-34
R & D program	ID-No
Subscription beginning (dd-mm-yy)	mm-dd-yy
Subscription end (dd-mm-yy)	mm-dd-yy
Version number	????

**User Name & Contact Details**

Subscribed	Ordering		
5	<a href="#">View</a>	<a href="#">Add</a>	<a href="#">Delete</a>





*Fig. 8. CMI Web page - Customer Control Panel*

USER:

Test plan was agreed.

### **25th February 1998 - Adding, Deleting**

USER:

At 14:05 CET: Zlatica initiated ADDITION of user: R.Zagolla zagolla@telekom.dbp.de DT Germany. The CMI indicated that the confirmation of the addition will be notified by e-mail what was not done by the CMI e-mail.

At 15:59 CET The notification of ADDITION of user was received through the ordinary e-mail from Paul telling that the ADDITION is evident (grey colored) but he didn't receive any notification by the CMI e-mail.

JAMES:

No (CMI) e-mails had been received regarding alterations to Paul's and Zlatica's details. Checked the CMI and at 15:45 CET found that R Zagolla had been added but all the alterations were still grey, and not accepted.

USER:

Between 14:00-16:00 CET: Zlatica tried to DELETE the user B but it was not allowed because the DELETION button for the specified user was not present.

At 14:41 CET: Zlatica sent e-mail 1 initiated through the button HELP with intention to ask JAMES which users can be deleted by whom. The e-mail address was "james@dashi.berkom.de.berkom.de".

At 14:47 CET: Zlatica received returned e-mail 1 with diagnostic "Host unknown".

At 15:52 CET: Zlatica sent the e-mail 2 to the Provider with changed address of e-mail 1 to "james@dashi.berkom.de" as it doesn't seem to be a correct RFC-822 address.

At 15:56 CET: Zlatica received indication of returned e-mail 2 with diagnostic "User unknown".

### **26th February 1998 - Adding, Deleting**

USER:

At 10:00CET: Zlatica sent the e-mail to the Provider informing him about the problem to delete the user B.

At 18:44 CET: Zlatica received e-mail from Paul where he suggested to try to delete the user F after modifying the connection B-A to A-F in the test 6.1 as it was clear that the user who has active connections is protected from deleting.

JAMES:

Still no (CMI) e-mails received. Discovered finally that the alterations were awaiting modification, hence reason why they were grey. No e-mail had alerted me to changes so I had no reason to check or authorize anything. Also not clear at all HOW to authorize

these alterations, but managed to find them eventually and very quickly authorized all the outstanding alterations.

Because no e-mails were being received (or sent?) I tried to use the facility embedded in the CMI.

Mail from Zlatica came saying that she could not delete user B. We suspected the reason was that it was involved in connections, no mention of this in the manual. It was proposed that Zlatica changed connection 2 to A-F and then try to delete user F to prove whether this is the case.

(it should be added here that included in this deleting section should be the denying of a request but there was little point in doing this since the e-mails were not in operation.)

### **27th February 1998 - Modifying**

USER:

At 11:31 CET: Zlatica performed MODIFY of cell rate to 7076 for CBR connection ID 2. (B-A) but in the same time she changed the connection end points field from B-A to A-F as indicated in 5.1, Indication 3.

At 11:52 CET: Zlatica performed MODIFY of the test date to 31/03/98, for connection ID 4 (VBR).

As the connection end points field was not planned to be changed (B-A), she didn't touch it in

MODIFY but it was resetted to default values i.e. to A-A.

At 14:19 CET: Zlatica received the e-mail (not CMI mail) from Paul with confirmation for all changes.

At 14:45 CET: Zlatica turned back the connection end points field for the connection ID 4 (VBR) to B-A from the default value A-A.

JAMES:

Still no (CMI) e-mails.

Checked the CMI on the off-chance and found that Zlatica had made the changes and I authorized them.

### **2nd March 1998 - Multiple changes**

JAMES:

Paul completed the following alterations:

1. added CBR service A-G, 02/03-31/03, 4717, 10-21, text went grey and said check for response in e-mail
2. deleted user H, Rheinhard Zagolla, text went grey
3. modified ATM access to JAMES POP on connection 4, date now ends 31/03/98, not 31/12/98
4. used mailing facility to see if Zlatica receives the mail

### **3rd March 1998 - Viewing**

USER:

At 10:05 CET: Zlatica received normal e-mail from Paul asking if it is possible to recognize the modifications which he did and to confirm them.

At 13:54 CET: Zlatica checked the modifications without receiving any indication by the CMI e-mail and sent normal e-mail to Paul approving the changes.

The changes were (coloured grey):

- a) The User Name in the table User Name and Contact details under H is not authorized. It is not clear if this User waits an authorization to be added or to be deleted.
- b) In the table Currently subscribed Services ATM CBR subscriptions are increased from 3 to 4 and a new connection is created between user A (J. Aronsteim - Oslo) and user G (P.Richmond - London) with following details: A-G, duration 02/03/98 31/03/98, CBR 4717 4717, period of day 10-21 .
- c) In the Table Currently subscribed Services IP over ATM not approved items for the User site under E where in all three tables (Host IP Details, ATM access to JAMES POP and in Alternative access to JAMES POP). In the table ATM access to JAMES POP (User Site=4, Connection Identifier=4) User noticed the change in the field End-Date: from 31-12-98 it is changed to 31-3-98.

JAMES:

Zlatica recognised 1 and 3 of the above. It was not clear to her if user H in 2 was to be added or deleted. Also she received no e-mail from the facility in 4.

USER:

The VIEWING was done through all test period from 25 February till 4 March 1998.

A list of comments concerning suggestions for improving the CMI as well as suggestions concerning the User manual were forwarded to the CMI development team on March 6, 1998. The principal comment was that the CMI e-mail notification didn't work and we didn't know why. After the tests were concluded, the people from WP4.2 explained (on March 5, 1998) that the CMI e-mail notification works only for the user A what was not evident from the available documentation on CMI.

In general, the conclusion was that the CMI is highly promising management interface, but it was still at the prototype level and still not applicable in the real life situations.

### **Test related problems and general suggestions**

One of the problems was the coordination of many participants in the tests (specially in the PVC configuration) as it was done at each edge switch locally. The tests of remote PVC configuration from the central Management Platform showed that this is already possible to configure PVCs using the standard MIB.

The delay on CMI tests caused the lack of more detailed test scenario and the lack of participation of CMI designers. Because of this we were not in the possibility to test how the CMI e-mail works (the most interesting part of the CMI).

The focus was at the ATM switches while the routers with ATM interfaces were not enough tested.

### **Relevance for service and migration suggestions**

The real management service was not offered by the JAMES which caused that we had to test the management possibilities only at the NRN edge devices. One exception was the CMI, but it is still only a prototype.

SNMP based management platforms on the user premises could be used for the M3 interface (Customer Network Management for ATM Public Network Service) but the security and access control is still unsolved problems. We expect that the SNMP version 3 will accelerate the solutions for these problems.

Also, TMN solutions are already entering at the market, not only at the PNO side but also at the user side. The TMN solutions have to be taken in account as an alternative to SNMP in the ATM technology.

One very promising ATM specific management concept is the concept of OAM flows, but it is not clear why the OAM flows are not fully standardized and implemented by the ATM switch vendors. OAM flows are mandatory and relatively simply to implement. They are totally independent of the management protocol (TMN or SNMP). Today, the OAM is still lack the relevant MIB extensions which could enable to build management applications for ATM-like services.

## Further studies

For the final report it will be very useful to include the studies status-of-the art in the domains of TMN and SNMP with the ATM technology in mind.

## Bibliography and references

1. Internet Draft, Ly, Noto, Smith, Tesink, Definitions of Supplemental Managed Objects for ATM Management, (draft-ietf-atommib-atm2-01.txt), July 1997
2. ATM Forum Specification: Customer Network Management for ATM Public Network Service (M3 Specification), 1996
3. ATM Forum Specification: UNI v. 3.1, 1995
4. ATM Forum Specification: UNI v. 4.0, 1996
5. ATM Forum Specification: ILMI v. 4.0, 1996
6. ATM Forum Specification: Introduction to ATM Forum Performance Benchmarking Specifications, 1996
7. ITU-T, I.610, Integrated Services Digital Network (ISDN), Maintenance Principles, B-ISDN Operation and Maintenance Principles and Functions, November 1995
8. TU-T, I.751, Integrated Services Digital Network (ISDN), B-ISDN Equipment Aspects, Asynchronous Transfer Mode, Management of Network Element View, March 1996
9. ITU-T Recommendation I.356, Overall network aspects and functions - Performance objectives, B-ISDN ATM layer cell transfer performance, October 1996
10. CCITT Recommendation M.20, Maintenance philosophy for telecommunications networks, 1992
11. CCITT Recommendation M.3010, Principles for a telecommunications management network, 1992
12. ITU-T Recommendation M.3610 - Principles for applying the TMN concept to the management of B-ISDN, May 1996
13. ATM Forum, Remote Monitoring MIB Extensions for ATM Networks AF-NM-TEST-0080.000, May 1997
14. IETF RFC 1695, M. Ahmed, K.Tesink, Definitions of Managed Object for ATM Management Version 8. using SMIV2.
15. CMI User Manual, JAMES WP4.2
16. Internet draft, Definitions of Tests for ATM Management, <draft-ietf-atommib-test-03.txt>

## Abbreviations

CER Cell Error Ratio

CES Circuit Emulation Service

CMI Customer Management Interface

IMA Inverse Multiplexer for ATM

SDU Service Data Unit

# Glossary

ABR	Available Bit Rate
ARP	Address Resolution Protocol
ATM	Asynchronous Transfer Mode
CBR	Continuous BitRate (ATM Forum: traffic class)
CER	Cell Error Ratio
CES	Circuit Emulation Service
CMI	Customer Management Interface
DCC	Data Country Code
DBR	Deterministic BitRate (ITU-T: traffic class, eq CBR)
E.164	(ITU-T addressing standard)
ICD	International Code Designator
IESG	The Internet Engineering Steering Group. Manages the working groups and standardization process in IETF
IETF	Internet Engineering Task Force ( <a href="http://www.ietf.org">http://www.ietf.org</a> ) The Internet protocol standardization body
ILMI	Interim Link Management Interface
IMA	Inverse Multiplexer for ATM
IP	Internet Protocol
ISO	International Standards Organisation
ITU	International Telecommunications Union
JAMES	A European experimental ATM-network.
LIS	Logical IP Subnetwork
MBS	Maximum Burst Size (ATM Forum: traffic parameter)

NHRP	Next Hop Resolution Protocol ( <a href="ftp://ds.internic.net/internet-drafts/draft-ietf-rolc-nhrp-11.txt">ftp://ds.internic.net/internet-drafts/draft-ietf-rolc-nhrp-11.txt</a> )
NHRP-R2R	NHRP for Destinations off the NBMA Subnetwork <a href="ftp://ietf.org/internet-drafts/draft-ietf-ion-r2r-nhrp-00.txt">ftp://ietf.org/internet-drafts/draft-ietf-ion-r2r-nhrp-00.txt</a>
NRN	National Research Network
NSAP	Network Service Access Point (OSI term)
OAM	Operations And Maintenance
P2MP	Point to Multipoint
PCR	Peak Cell Rate (ATM Forum: traffic parameter)
PIM SM	Protocol Independent Multicast, Sparse Mode
P-NNI	Private Network to Network Interface
PNO	Public Network Operator
PVC	Permanent Virtual Circuit
PVPC	Permanent Virtual Path Connection
RSVP	Resource ReSerVation Protocol Version 1 Functional Specification - Internet draft; <a href="http://www.internic.net/internet-drafts/draft-ietf-rsvp-spec-12.txt">http://www.internic.net/internet-drafts/draft-ietf-rsvp-spec-12.txt</a>
SBR	Statistical BitRate (ITU-T: traffic class, eq VBR)
SCR	Sustainable BitRate (ATM Forum: traffic parameter)
SDU	Service Data Unit
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

---

[Back](#) to table of contents