Video and Audio Streams Over an IP/ATM Wide Area Network

Mark McCutcheon Mabo Robert Ito Gerald W. Neufeld

UBC TEVIA Project
Transport **E**ncoded **V**ideo over **I**P/**A**TM
Technical Report 97–3
June 1997

University of British Columbia
Department of Computer Science

Technical Report 97–3
June 1997

Abstract

This is a survey of the state of the art in delivering IP services over ATM networks, as it stands in the second quarter of 1997. It also includes a look at the alternatives to that set of technologies. The technology and the choices are changing "on the fly", and have evolved significantly during the course of this project. Moreover, the issues are not exclusively technical, but in many respects reflect the great schism in the data communications world: connection-oriented *versus* connectionless networks. We have tried to present the technical issues and solutions along with an unbiased overview of the more "philosophical" issues. We indicate how we think the technology and the installed base of equipment is going to develop over the next few years, in order to give a picture of the future of ATM in data networking.

Contents

List of Figures

List of Tables

1 Introduction

We have attempted to provide an accurate reflection of the state–of–art for delivery of IP services over ATM, and the alternatives to that set of technologies, as of the time of publication. The technology and the choices are changing "on the fly", and have evolved significantly during the course of this project. Moreover, the issues are not exclusively technical, but in many respects reflect a perpetually rancorous schism. In this report we have tried to present the technical issues and solutions accurately, to give an unbiased overview of the more "philosophical" issues, and finally to indicate how we think the technology and the installed base of equipment is going to develop over the next few years. Clearly, the certainty with which we can present each of these overviews declines as we move away from looking at today's technology.

The great schism in the data communications world is: Connected or Connectionless? [1] It has been exacerbated by the widespread deployment of ATM for data backbones by telcos. Into this divide are subsumed any number of related conflicts:

- packet vs. cell
- L3 routing vs. L2 switching
- "best–effort" vs. resource reservation
- hop–by–hop forwarding vs. source routing
- soft state vs. hard state
- broadcast/multicast vs. point–to–multipoint
- receiver–driven multicast vs. sender–driven multicast
- connections over connectionless networks vs. connectionless over connection-oriented networks

Participants in the debate have taken extreme positions. According to the pure *(Connectionless)* view, there is no need or excuse for connection–oriented network technology. ATM deployment has peaked and is beginning to decline. It will be universally replaced by extremely fast IP routers, connected in a point–to–point mesh by high–rate SONET links (or Frame Relay, SMDS, or other packet data, non–switched technology). Some hold that this will extend into the PSTN, with

[1] Versions of this debate have existed for over twenty–five years, ever since connectionless packet switching was first defined.

even traditional telephone service moving to connectionless delivery. The 25 year history of the Internet has proved a golden design rule for robust internetworking systems: keep the network simple and put the complexity in end systems.

The contrary view, of course *(Connection–Oriented)*, is that ATM and its native protocols are sufficient for all communication needs, from voice and video to computer Wide– and Local–Area Networks and data communications. IP is just another protocol which will be carried on an ubiquitous ATM network, which will extend from Service Providers right to the desktop.

The ATM–centric view is quite unlikely to be realized in the foreseeable future. On the other hand, it's far less seriously proposed or widely advanced than the IP–centric view. Most large datacom corporations find it in their interest to develop a synthesis of these views, fielding a variety of solutions for carrying IP traffic over ATM/SONET physical networks. However, there is a large and serious camp in the IP–centric world which concentrates on defining and developing alternate link technologies. In this world view, the link technology should be unimportant, as long as it is fast, efficient, ubiquitous, and dumb. The key is extremely fast address lookups, high packet forwarding rates, and very high–bandwidth router backplanes, whether switched or bussed.

What are the primary strengths and weaknesses of connectionless IP *versus* connection–oriented IP–over–ATM networks? It is widely believed that packet switching provides more efficient usage of resources than circuit switching, and works best when carrying bursty regular data traffic. This truism is complicated by the discovery of "self–similarity" in network traffic behaviour at different levels of aggregation [69]. Much work remains to be done to clarify how this affects statistical multiplexing. On the other hand, circuit switching has the capability of providing hard Quality of Service (QoS) guarantees. But the most QoS–sensitive traffic — voice data — represents a decreasing fraction of the overall traffic in all types of networks. Live video is a most demanding application, in terms of bandwidth as well as other parameters which go to make up quality of service. Table 1 is a loose comparison of the parameters considered by IP Integrated Services and ATM in defining "Quality of Service".

Full–motion video is not possible across Internet connections at present, due both to lack of sufficient bandwidth and possibly the immaturity of resource reservation schemes. It is difficult to support hard QoS guarantees on a connectionless network, and there may be applications which require them; however, this is not the usual situation.

The question we want to answer isn't which approach is better, however, but whether and how IP and ATM will cooperate in the IP–dominant environment of the Internet? The rest of this report contains discussions of the various methods that have been developed or proposed for bringing IP and ATM together, and for

Table 1: Internet vs. ATM Quality of Service Parameters

INTERNET QOS [20]		ATM QOS [16]	
Packet delay	real–time (bounded)	Cell transfer delay	
	elastic (best–effort)	Cell delay variation	
Link sharing	assigned shares	Cell rate	peak
	traffic control		sustainable
Packet dropping		Cell loss ratio	
		Cell error ratio	
		Cell misinsertion rate	

making this combination a useful carrier for desired services.

2 Supported Services

In this report, we have considered those delivery mechanisms which are suited (or at least proposed) for the broadest range of digital multimedia. In this context we refer to some combination of digital video, digital audio, and both static and active presentational displays (e.g. lecture slides or "whiteboards"). Some general categories are:

1. Entertainment multimedia
   - (a) motion picture and television production
   - (b) pay–per–view live broadcast
   - (c) video–on–demand
   - (d) interactive gaming
   - (e) internet telephony
2. Business/commercial multimedia
   - (a) enterprise–wide presentations
   - (b) multimedia conferencing
   - (c) enterprise–wide telephony
   - (d) workgroup collaboration
   - (e) interactive sales
3. Medical multimedia

   (a) medical real–time imaging
   (b) medical image storage and retrieval
   (c) remote consultation and diagnosis
   (d) interactive “rounds”

4. Educational and scientific multimedia
   (a) distance education
   (b) multimedia libraries
   (c) virtual conferencing
   (d) scientific presentations
   (e) committee working sessions (e.g. IETF meetings)

It is clear that these categories are not all–inclusive, nor are they mutually exclusive. They form a framework to focus our thinking. It is also interesting to note that only in the case of entertainment multimedia is the service the end–product being sold. Hence expectations for quality and availability are likely to be higher here than in many of the ostensibly more “serious” applications, with the exception of medical technology.

These applications make use of a range of encodings with vastly different delivery requirements, in terms of bandwidth, delay sensitivity, and the other elements of the Quality of Service concept. Table 2 summarizes a number of leading video and multimedia encodings, in order to compare their modes of delivery and required data rates. Table 3 does the same for exclusively audio encodings.

The protocol stacks which use these encodings and are employed by high–level applications are the proper subject of this report. Layerings of protocol stacks permit audio and video to be encapsulated and delivered to viewers or recipients across a very wide range of underlying media and under an equally broad range of expectations of quality and cost. They involve unicast, multicast, and broadcast models; high quality, guaranteed delivery service and slow, low quality links. In Section 5 we present an overview of these protocol stacks.

The most basic (and most contentious) issue which arises in any consideration of multimedia is the concept of Integrated Services (**IS**). RFC1633 [20] provides an overview of Integrated Services in the Internet. The term refers to an Internet service model which includes “best effort” service, real–time service, and controlled link sharing. By “controlled link sharing” we understand a situation where traffic is segregated into administrative classes, each of which may be assigned some minimum fraction of link bandwidth under overload conditions. In other words, this is the most basic statement of a Quality of Service model.

Table 2: Video and Multimedia Encodings

Standard or Format	Service	Comp. Ratio	Bandwidth Range	Source
YUV/CIF	vic	5:1	37 Mbps (320X240px, 30fps)	[103], [95]
JPEG	vic		1.35 Mbps (320X240px, 30fps)	[103]
NTSC	TV	n.a.	209.5 Mbps (600X485px, 30fps)	[113]
PAL	TV	n.a.	400.2 Mbps (580X575px, 50fps)	[113]
H.320/H.261	vid.conf.	24:1	64 kbps–2 Mbps	[113]
	RealVideo 1.0		28.8 kbps (newscast) 55 kbps (full motion) 100 kbps (“near” NTSC)	[29]
DVI		160:1	1.2–1.5 Mbps	[113]
CDI		100:1	1.2–1.5 Mbps	[113]
M-JPEG	CMFS	7:1 27:1	10 Mbps 20 Mbps	[113]
MPEG–1	storage	100:1	1.2–2.0 Mbps (352x240px, 30fps)	[113]
MPEG–2	any	30:1 ... ... 100:1	4–60 Mbps 1.5 Mbps (VHS, 352x240px) 5–6 Mbps (b’cast, 1440x1152px) 7 Mbps (studio, 1920x1080px)	[113], [94], [95]
MPEG–4			4.8–64 kbps (176X144px, 10fps)	[94]
ITU-R 723		3:1 5:1	32 Mbps 45 Mbps	[113]
ITU-R 601	TV	n.a.	140–270 Mbps (320x480px)	[113], [95]

Table 3: Audio Encodings

Standard or Format	Service	Comp. Ratio	Bandwidth Range	Source
G.728	telephony	4:1	16 kbps	[95]
G.726	telephony	1.6:1	16 kbps	[95]
		4:1	40 ibps	
G.722	telephony	3.5:1		[95]
ISO audio	MPEG			
Layer 1	Philips DCC	7.8:1	192 kbps	[94],
Layer 2	MUSICAM	11.7:1	128 kbps	[95]
Layer 3	ASPEC	23.4:1	64 kbps	

The existing Internet at present offers only the "best effort" component. RFC-1633 presents IS as an *extended service model* to the basic architecture. It lists three arguments which have been raised against implementing an IS model in the Internet:

- "Bandwidth will be infinite" [2]
- "Simple priority is sufficient" [3]
- "Applications can adapt" [4]

and briefly summarizes the reasons these do not constitute an adequate substitute for resource guarantees in the Internet. RFC1633 is now 3 years old, and "best effort" service on the Internet has only continued to degrade in the interim. Although all multimedia applications have been designed with this type of "best effort" delivery as default, two powerful elements have combined to worsen the situation intolerably:

1. exponentially growing demand for Internet access
2. insufficient resources even for existing traffic

It is at least debatable whether progress in network speed alone can keep sufficiently head of network ubiquity (hence traffic growth) to relieve any need for resource guarantees. Whether these need to be hard guarantees, statistical guarantees, or susceptible to some other means of predicting access to services is hardly close to resolution.

3 The Standards Process: Protocol Stacks

Three standardization bodies in part share and in part compete in shaping the development of Integrated Services over ATM. These are the Internet Engineering Task Force (**IETF**), the International Telecommunications Union - Telecommunication Standardization Section (**ITU-T**), and the ATM Forum (**ATM-F**).

The Internet Engineering Task Force is concerned primarily with the upper layers of the internet hierarchy, dealing with particular physical network layers only as required. It is the protocol engineering and development body for the Internet, under the aegis of the Internet Architecture Board (IAB). It is a open international

[2]OC-192 and Gigabit routers, to start
[3]e.g. IPv4 TOS bits plus charging for priority service
[4]e.g. Jacobson VAT model

community of network engineers and researchers who are concerned with the evolution of the Internet protocols and architecture. The IETF has a number of Working Groups, each addressing a particular issue (see Appendix A). Much of the work is done in various mailing lists. IETF also holds three meetings every year. Discussion in the mailing lists and meetings will result in the production of Internet Drafts, which have a lifetime of 6 months. Drafts will either be discarded, updated, or further refined into RFCs. We are aware of drafts which have gone through as many as 13 different releases and have not yet made it to the stage of an RFC [21]. The working groups related to IP over ATM and real-time traffic support are:

avt Audio/Video Transport

bmwg Benchmarking Methodology

idmr Inter-Domain Multicast Routing

idr Inter-Domain Routing

intserv Integrated Services

ion Internetworking Over NBMA (merger of "IP over ATM" (ipatm) and "Routing Over Large Clouds" (rolc)

ipcdn IP over Cable Data Network

ipngwg IPNG

issll Integrated Services over Specific Link Layers

mboned MBONE Deployment

mmusic Multiparty Multimedia Session Control

mospf Multicast Extensions to OSPF

mpls Multiprotocol Label Switching

ngtrans New Generation Transition

nimrod New Internet Routing and Addressing Architecture

ospf Open Shortest Path First IGP

qosr QoS Routing

rip Routing Information Protocol

rsvp Resource Reservation Setup Protocol

rtfm Realtime Traffic Flow Measurement

The Drafts and RFCs emerging from the IETF are public documents and freely available from a number of sources, indicating its heritage as a predominantly North American, government–funded body.

The ITU-T approaches networking from the opposite direction, building upward from network hardware. It is responsible for the definition and standardization of SONET, SDH, and the B-ISDN which is the origin of ATM. It is an organization dominated by telecommunications interests, as the IETF is dominated by datacom interests. At present, the ITU-T has 14 "Study Groups" (see Appendix B), of which the work of 8 is directly relevant to our interests:

Study Group 2 Network and service operation

Study Group 7 Data networks and open system communications

Study Group 9 Television and sound transmission

Study Group 11 Signaling requirements and protocols

Study Group 12 End-to-end transmission performance of networks and terminals

Study Group 13 General network aspects including GII

Study Group 15 Transport networks, systems and equipment

Study Group 16 Multimedia services and systems

The ITU-T documents are published and distributed commercially, at significant cost. It is self–funding as might be expected of an organization with truly international membership.

The ATM Forum is an industry-supported body concentrating on ATM related protocols. It is a world-wide organization explicitly aimed at promoting the use of ATM, although it also functions as a *de facto* standardization body. The ATM Forum has 14 Technical Working Groups under its Technical Committee at present. The Technical Groups related to IP over ATM are:

B-ICI B-ISDN Inter-Carrier Interface

LANE LAN Emulation

MPOA Multi-Protocol Over ATM

PHY Physical Layer

PNNI Private Network-Network Interface

RBB Residential Broadband

TM Traffic Management

. Network Management

. Service Aspects and Applications

. Signaling

. Testing

ATM Forum documents, like those of the ITU-T, are sold commercially, and in some instances published in book form by commercial publishing houses (e.g. UNI 3.1). However, the working documents are available from the ATM Forum's Web site.

4 Issues Related to IP over ATM

There are a number of issues that need to be addressed in any IP over ATM solution. These include Virtual Path–Virtual Circuit (VP/VC) usage and circuit setup–teardown policy, bandwidth efficiency and encapsulation, address support, routing/address resolution, multicast, QoS support, signaling, and IPv6 support. The overall determinant of success will be *scalability*.

All of the native modes of ATM as a sublayer of the B–ISDN have been developed with the aim of achieving an unprecedented degree of scalability. The very large ATM address space, whether NSAP end–point identifier or E.164, was defined partly for this reason, as was the P–NNI with its undeniable complexity and its multiple hierarchical layers. However, this scalability is dependent upon re–use of the VP/VC space at each network interface by employing Switched Virtual Path/Virtual Circuits (SVP/SVC), and hence on the speed with which signaling can set up and tear down circuits on demand.

The VP/VC usage problem in IP over ATM arises from the fundamental difference between classical IP and ATM – the connectionless vs. connection–oriented approach. In a network where all traffic flows across connections, and the lifetime of a connection is considerably greater than the time required to set it up and tear it down by end–to–end signaling, SVP/SVCs are reasonable and their use is indeed scalable. However, a connectionless network has very different requirements. Each packet (or frame, or Protocol Data Unit) is a relatively small and self–contained package of information, which is expected to make its own way across the network. As it is, at least in principle, unrelated to the packet which precedes it or follows it, its characteristic timescale is very short. It is untenable to imagine setting up an end–to–end connection, with all the signaling and state information which must be delivered to switches along the way, to route a single packet across even a relatively small network, and such a process is manifestly unscalable in a global Internet. But the connectionless paradigm is the very soul of the Internet. In the broader context, we would be justified in mapping "flows" over IP to Virtual Circuits. However, traffic analysis shows that by reasonable definitions, flows average only around 100 packets. This sets fairly stringent limits on how much time can usefully be spent recognizing flows, and in setting up Virtual Circuits or other stateful paths across the Internet to carry them.

There is no simple solution to this problem. In current ATM implementations, cell switching can be done very quickly but connection establishment is much slower, usually by several orders of magnitude. For instance, the StrataCom BPX^{tm} Service Node (ATM switch), which supports 384,000 active connections per node with a throughput of 20Gbps, processes only 4000 "calls" per second [5]. This results in a difficult decision between using persistent VCs, either Permanent Virtual Circuits or "Soft PVCs" [6], and setting up new VCs on demand. There are pure "connectionless" approaches, where VCs are only used to connect neighboring routers, and pure connection-oriented approaches, where new VCs are always created. Both extremes have disadvantages; either they do not make full use of ATM capability or they are too expensive, or both. It is commonly agreed that there are a spectrum of possibilities in between. The ability to scale from local network to Internet proportions is a major issue.

On the other hand, connections are frequently necessary even in a connectionless network. TCP is the most common means of achieving connections in the Internet, but only the end–stations know about such connections in current implementations. This doesn't help with any type of service guarantee, bandwidth reservation, etc., which requires setting up state in intermediate routers or other switchpoints. As we discussed in the Introduction, there is a significant and growing de-

[5] *StrataCom BPX Service Node,* ©Cisco Systems 1988–1997

[6] "Soft PVC" refers to the situation where a Permanent Virtual Circuit is set up across both source and destination UNIs, but P–NNI is used to connect these links across all intervening NNIs with normal, though persistent, Switched Virtual Circuits. The intention is to support end–systems which are not capable of signaling.

mand for service guarantees for the delivery of Integrated Services, and the most straightforward method of enabling a service guarantee is by offering connections. A relatively non–intrusive method of establishing end–to–end connections even in a purely IP environment is through flow identification. This is typically done by recognizing some combination of source IP address and port number, destination IP address and port number, and protocol type. Cisco's NetFlow Switch Software^{tm}, an upgrade to their high–end 7000 and 7500–class routers' IOS^{tm} software, does just this kind of flow recognition, coupled with simplified next–hop lookup. An extremely sophisticated solution, but one which has yet to see any deployment in the Internet, is IPv5, otherwise known as ST2 [47, 89, 109]. ST2+ actually turns the Internet into a connection-oriented network.

Cisco's NetFlow sets up flow state in the router on the first packet of a potential flow, which can lead to an actual increase in the router's burden with no concomitant performance improvement under some traffic patterns. Statistical approaches have been proposed in which the border router will set up a circuit only after a number of packets greater than some threshold have been sent in a recognizable "flow". This raises questions of the speed and efficacy of flow identification. As we will see, the same issue arises in "Multi–Protocol Label Switching", the current generic term for coupled IP–routers and ATM switches. Algorithms or heuristics also need to be devised for closing connections and tearing down circuits. The problem is not completely solved. If the number of packets that must be seen before setting up a new circuit is too high and if the tear–down condition is too tight, there will still be inefficient use of network resources. Better solutions to this problem require QoS information to be used when making the decision. This requires QoS routing, which by itself is very much an open issue.

Another class of fundamental questions revolve around the extent to which it is beneficial to map different flows above IP into separate connections below IP, or conversely to merge different flows with the same destination into a single connection below IP. This will be discussed further along with MPLS, Section 5.7.

Encapsulation deals with the way in which IP packets are sent over ATM VCs. RFC1483 [60], the basic specification on multiprotocol encapsulation over ATM AAL5, defines 2 techniques: LLC/SNAP encapsulation, which allows different protocols to be multiplexed onto a single VC; and VC–based multiplexing, which assigns each protocol a separate VC. The choice of which to use clearly depends upon a trade–off between VC conservation and encapsulation efficiency.

Routing/address resolution is the central theme of much ongoing research in the IP over ATM domain. Different models for IP/ATM integration provide very different answers. We will go over these solutions in considerable detail later, especially the scalability implications for Internet deployment. I–PNNI is the most thoroughly ATM–centric approach; the others are generally ways of making an ATM network look like a classical broadcast network to applications.

The Multicast Address Resolution Server (MARS) model is the most widely adopted method of supporting multicast over ATM. Special consideration is also given to Protocol–Independent Multicast – Sparse Mode (PIM–SM) over ATM. Section 5.8 reviews this solution.

RSVP is a resource reservation protocol used to support QoS guarantees for data streams on the Internet. With RSVP, many other service types besides "best-effort delivery" might be supported on the Internet; for example, real-time delivery service. ATM has been designed with native support for Quality of Service guarantees, and might seem a perfect substrate for implementing Integrated Services for IP. However, the fit between IP RSVP and ATM QoS may not be as easy as at first appears. We discuss the problems in Section 5.9.

Signaling for call setup and teardown and other administrative functions is quite different under ATM than in more common network media. Older ATM switches generally supported only out–of–band signaling, but these are obsolete and have largely been replaced. Signaling in true ATM networks is done using OAM [7] cells, sometimes over well–known PVP/PVCs, sometimes within application–level Virtual Circuits. Older proprietary signaling schemes are still in use (primarily Fore Systems' SPANS) but are gradually being displaced by industry–standards for UNI and NNI signaling.

The final major issue we will discuss is IPv6 support (Section 5.10). IPv6 is the "Next Generation" IP protocol, designed to salvage the Internet from an address space explosion. Many new features have also been added to IPv6, such as support for flow labeling, which is a good thing for real-time traffic, and neighbor discovery, which is difficult for ATM to handle, since it assumes the underlying network inherently supports multicast.

5 Techniques of IP Over ATM

There is a plethora of models for running IP over ATM, some rooted in the IETF, some in the ATM Forum, and some developed by specific data communications companies. We will review the following techniques:

1. LAN Emulation (LANE)

2. Classical Model IP over ATM (RFC1577)

3. Next Hop Resolution Protocol (NHRP)

[7] Operations, Administration and Maintenance

4. Multi–Protocol Over ATM (MPOA)

5. Integrated PNNI (I–PNNI)

6. Multi–Protocol Label Switching (MPLS)

7. Multicast over ATM

8. IP Integrated Services over ATM

9. IPv6 over ATM

These are not all equivalent in function or purpose; for instance, LANE and RFC1577 really are unique methods of running IP as an overlay on a normal ATM network; MPLS uses ATM hardware in a non–ATM fashion; and I–PNNI is an attempt to integrate routing across ATM networks and legacy, connectionless networks. Nevertheless, these are the major areas of activity for IP–ATM integration at present, and thus within the mandate of this report. We will look first at specific methods for integrating IP with ATM (items 1 – 6 above), then see how they meet the challenges defined in items 7 – 9.

5.1 Signaling

Before discussing techniques for delivering IP service over ATM, we will look briefly at signaling. There are a limited number of ATM signaling technologies, and these are used in many of the IP–over–ATM schemes; MPLS models form the main exception, using Internet Layer 3 protocols for all signaling and database distribution. We will briefly discuss Fore Systems' proprietary SPANS signaling, then the ATM Forum UNI signaling, which is based on the ITU Q series recommendations [85].

Simple Protocol for ATM Network Signaling (SPANS)

ATM Forum standards have been slower than anticipated in maturing. We will discuss P–NNI in a following section. It is sufficient at present to note that P–NNI Phase 1 reached standardization only in April 1996 and is not widely implemented; its predecessor, P–NNI Phase 0, otherwise knows as the Interim Inter–Switch Signaling Protocol (IISP), is a very limited subset of the NNI protocol, supporting only manually configured static routes and minimal signaling. Fore Systems introduced SPANS, a proprietary UNI and NNI protocol, which is UNI 3.0–compliant. The NNI component permits the exchange of NSAP routing information to enable automated construction of a full–mesh PVC connection network between switches in a cloud. The UNI component allows a source device to establish SVCs across that cloud, using signaling running over the PVC mesh, for actual data transfer. According to Fore Systems [8] SPANS provides an IP tunneling capability across an ATM network, though whether this exists independently of their LANE or Classical IP implementations is unclear.

UNI 3.0 and UNI 3.1

UNI 3.0–compliant TAXI 100Mbps switch line and host interfaces have implemented various proprietary connection–control signaling schemes, for instance IBM's XON-XOFF rate control protocol [102]. However, the standard signaling system for UNI 3.0 and UNI 3.1 is a subset of the capabilities of the ITU–T Q series recommendations. Q.93B and Q.931 are Narrow–Band ISDN signaling protocols (DSS–1) and were the original models used for UNI 3.0. However, they provide no support for Virtual Paths or Virtual Channels, so the DSS–2 recommendation's Q.2931 was quickly adopted by the ATM Forum. It has been extended to provide support for multicast (Q.298x Point–To–Multipoint) and for implicit QoS parameter negotiation.

DSS (Digital Signaling System) UNI signaling is a Layer 3 responsibility, and the UNI has its own Layer 3 implementation quite independent of any Layer 3 data transport capability, such as IP. UNI 3.0 made use of an early ITU–T data link protocol set known as Q.SAAL (Signaling ATM Adaptation Layer). This is not compatible with the Q.2931 data link protocol, SSCOP (Service Specific Connection–Oriented Protocol), a TCP–like reliable delivery protocol [61]. Hence UNI 3.0 and UNI 3.1 are similar in function but not interoperable.

UNI 3.0/3.1 signaling [32, 33] provides for connection setup, teardown, and status inquiry. The Point–to–Multipoint extensions provide for root–initiated multicast trees. A very detailed, though now slightly dated, survey of UNI signaling systems due to Stiller [108] covers ATM Forum UNI specifications through UNI 3.1.

UNI 4.0

UNI 4.0, like P–NNI Phase 1, was completed in April 1996 [34]. It adds a set of optional capabilities to the basic set embodied in UNI 3.0/3.1. Of considerable importance to IP–over–ATM efforts in the long run is support for "Leaf–Initiated–Join" of Point–To–Multipoint groups, the model for all current IP multicast efforts. At present, however, LIJ is not supported by native ATM routing protocols, nor used by such IP–over–ATM models as LANE and MPOA. Of more immediate interest,

[8] FORE: IP Tutorial, `<http://www.fore.com/atm-edu/tutorial/spans3.html>`

UNI 4.0 signaling provides for explicitly signaled QoS parameters, proxy signaling [9], "anycast" addressing for advertising well–known services, and Available Bit Rate (ABR) service. UNI 4.0 deployment has been gradual, with many vendors implementing only parts of the standard in current switches (ABR signaling and anycast in particular).

5.2 LAN Emulation (LANE)

LAN Emulation differs from all other IP–over–ATM schemes described in this report in that it uses ATM as a MAC–level protocol below LLC, while the others use ATM as a data link protocol below IP. It uses an *overlay* model to run transparently across existing ATM switches and signaling protocols; it is in essence a protocol for bridging across ATM. It makes no pretense of being an internetworking protocol, nor of dealing with the issues of scalability this would involve. It is solely a Local Area Network protocol, like Ethernet.

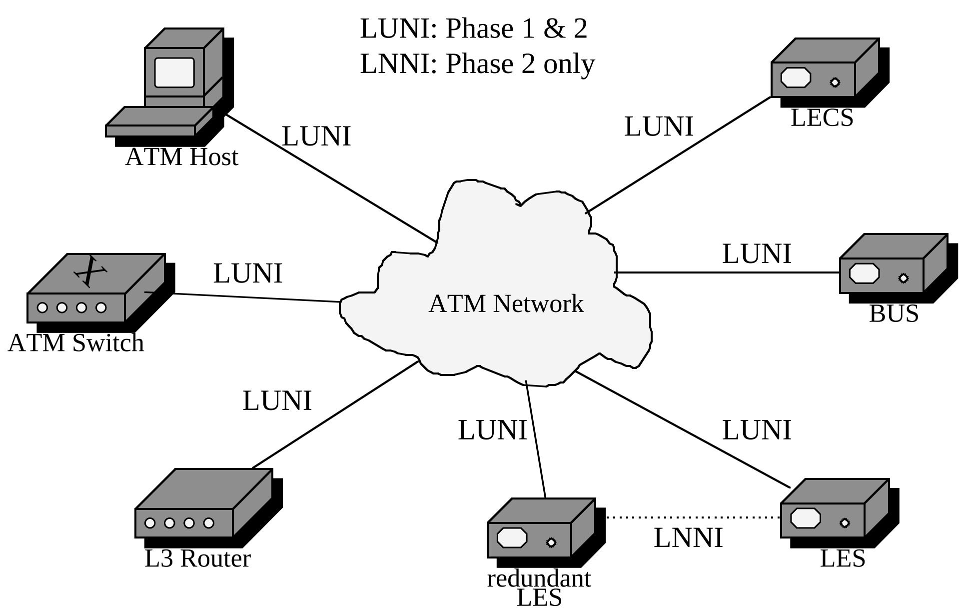


Figure 1: LAN Emulation Components and Protocol Interfaces

LANE divides an ATM network into multiple emulated LANs. These LANs operate independently and communication between emulated LANs is only possi-

[9]Proxy signaling allows a third node to set up connections between two other nodes, and is intended for use in ATM–based Residential Broadband systems

ble through routers or bridges. Figure 1 indicate the relationship between entities in LANE. Hosts in an emulated LAN are called LAN Emulation Clients (LEC). Each emulated LAN has a LAN Emulation Server (LES), a Configuration Server (LECS), and a Broadcast and Unknown Server (BUS). The Configuration Server assigns hosts in an ATM network to different LANs, the Broadcast and Unknown server handles all the broadcast/multicast traffic, while the LES is responsible for the LAN Emulation Address Resolution Protocol (LE_ARP). LE_ARP allows the LES to fulfill the basic responsibility of LANE, resolving MAC addresses into ATM addresses. This allows LECs to set up direct SVC connections between themselves for unicast data forwarding. Broadcast/multicast traffic is sent first to the Broadcast/Unknown server and then redistributed to all the receivers. ATM addressing schemes are flexible. Connection from the LES to the LECS may occur across a Permanent Virtual Circuit (PVC), be initiated from a "well-known address", or by using a protocol defined in the Integrated Local Management Interface (ILMI). A web of "permanent" switched virtual circuits, both bidirectional and unidirectional, are used to interconnect the LECs, the LECS, the LES and BUS for signaling and control. Figure 2 illustrates the relationships between LEC, LES, LECS and BUS.

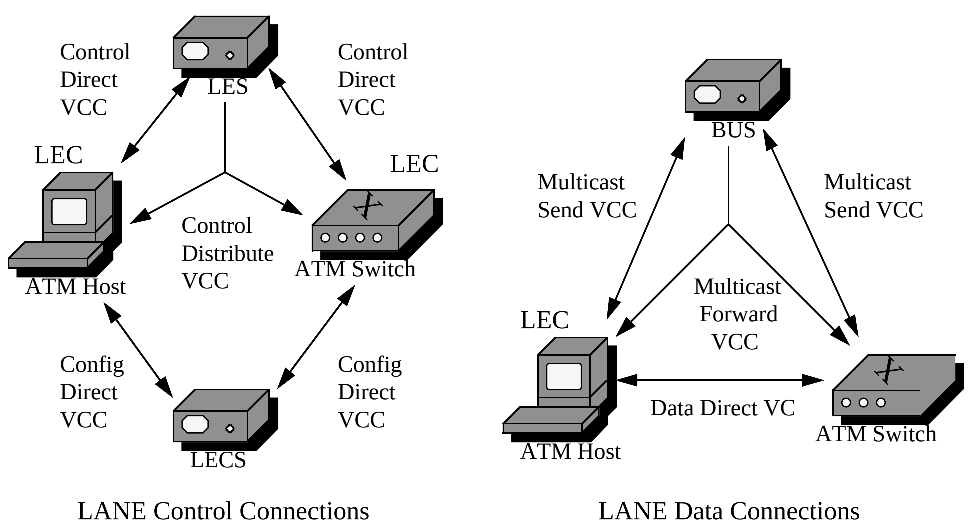


Figure 2: LAN Emulation Control and Data Connections

The LAN Emulation protocol defines mechanisms for emulating either an Ethernet (802.3) or Token Ring (802.5) LAN to attached host LECs. Supporting IP over LAN Emulation is the same as supporting IP over either of these IEEE 802 LANs, with no modification to higher–layer protocols such as the common NDIS driver interface for IP and similar protocol stacks. It should be noted, however, that LANE provides no means of directly connecting between Ethernet and Token Ring emulations. A gateway is still required to bridge between them. Forwarding packets between different emulated LANs must be accomplished via routers, either ATM-attached conventional routers or a form of ATM router implementing LANE at two or more interfaces to different emulated networks.

LANE Phase 1 is currently deployed, while Phase 2 is not yet complete. Phase 2 will add to the LAN Emulation NNI (LNNI) protocol to permit redundant LESs and replicated BUSs, as well as allowing for hierarchical BUSs. This is intended to avoid both single point failure modes and performance bottlenecks. Vendors such as Fore Systems have introduced redundant LES and BUS, and LEC failover function, in their current "Phase 1" releases in anticipation of the standard.

5.3 Classical Model (RFC1577)

The so–called "Classical IP" model is described in RFC1577 [66]. A new IETF draft, "Classical IP and ARP over ATM" [68], will obsolete RFC1577 when it becomes an Internet standard.

In the Classical model, the conventional IP subnet architecture is preserved. ATM adapters are treated as a network interface to the IP protocol stack. ATM networks under this model are divided into Logical IP Subnets (LIS) in which all the members have the same IP network/subnetwork address and netmask. Each member is connected to the ATM network directly and should be able to communicate with other members in the same LIS directly via ATM (that is, a full mesh of VCs is established among members of the LIS). Each member should also be able to map between IP addresses and ATM NSAP–format addresses using an ATM–based ARP and Inverse ARP service – ATMARP and InATMARP. One or more ATMARP/InATMARP servers may be used to provide address resolution in a unicast ATM environment for all members in the LIS. Figure 3 illustrates the manner in which ATMARP and InATMARP servers function to map back and forth between NSAP–format ATM addresses and IP addresses.

RFC1483 [60] encapsulation is used in Classical IP. LLC/SNAP multiprotocol encapsulation is used for transmitting IP data packets, while single–protocol VC Multiplexing is used for OAM functions. Routing architecture in the IP network remains unchanged. Traffic across LIS boundaries must be forwarded by a router which is a member of both LISs even though it might be physically possible to establish a direct VC connection between the source and destination (i.e. they have a physical connection at the ATM level).

This model has been proven to be quite successful when the number of nodes in the ATM network are not too large. Most of the current IP/ATM networks, such

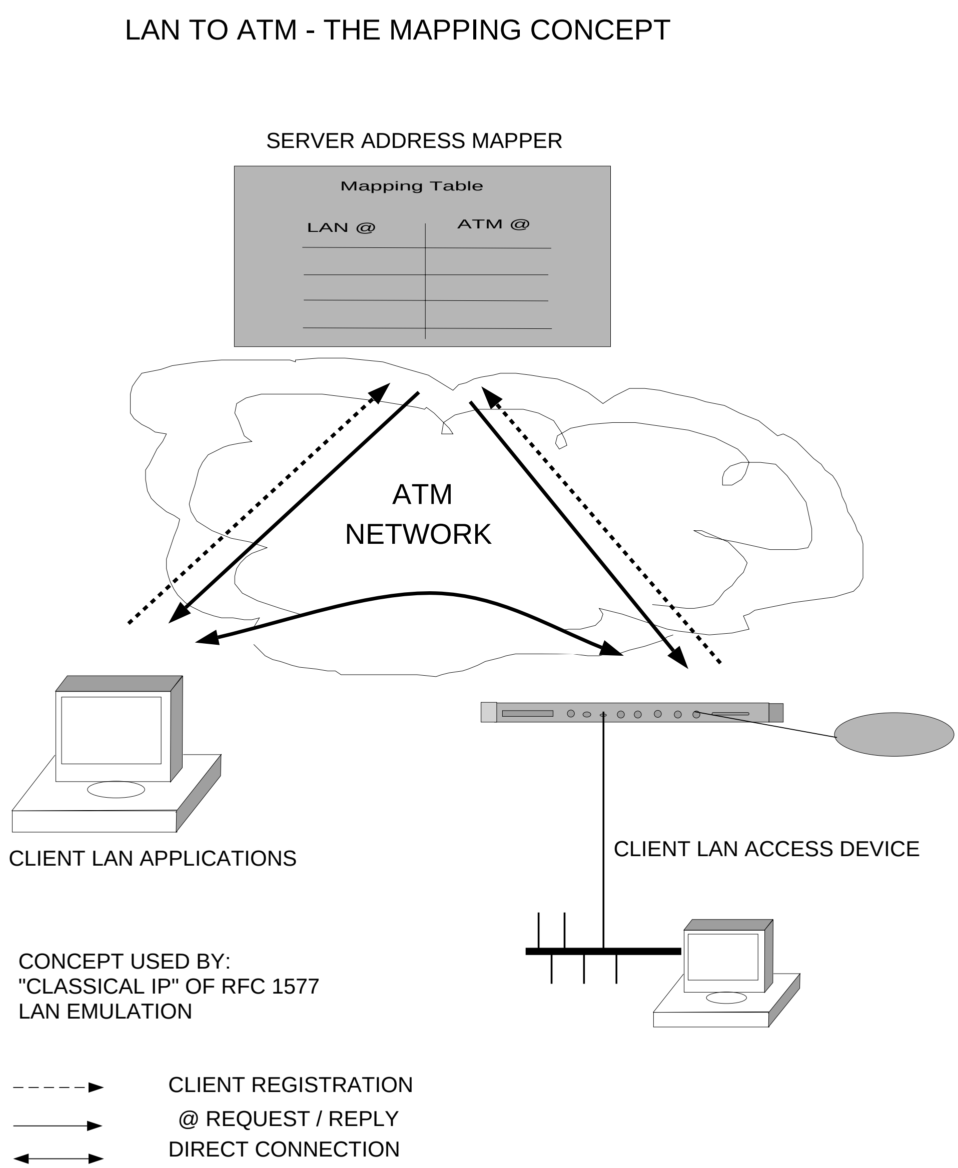


Figure 3: LAN to ATM Mapping Concept ©IBM Corp. 1996

as the CANARIE National Test Network, use Classical IP.

A problem with this model is that there may be unnecessary hops when the source and destination hosts are in different LIS, within the same ATM cloud. This is not an accident; it is called “Classical Model” because it follows RFC1122’s [19] *Requirement for Internet Hosts – Communication Layers* that any packet for a destination outside the source node’s IP subnet **must** be sent to a default router; this requirement is a significant limitation on the functionality of ATM clouds, however. Consider the following example:

```
Source--S1----S2-----S3-----Destination
                     |
                     R
```

Here we have three ATM switches: S1, S2, and S3. Source, S1, S2 and R belong to LIS1, Destination, R, S2, S3 belong to LIS2. Traffic from Source to Destination would follow the path Source–S1–S2–R–S2–S3–Destination. Obviously the shortest path should be a direct VC between the Source and the Destination, but this is not permitted by the Classical Model.

There are some limitations on the size of the LIS imposed by IP routing functions. The extra hop problem becomes much worse when the size of the ATM cloud and the number of LISs are very large. Furthermore, the necessity for IP processing at every router may add greatly to the latency experienced by data flows, at least with conventional IP routers.

Both LANE and the Classical model have the limitation of not using ATM’s ability to set up direct VCs across virtual LAN boundaries, illustrated in Figure 4. In some circumstances this is considered to be a virtue; ELANs under different administrative control are logically quite separate even when individual switches are physically connected. This may be a security advantage. The biggest advantage of LANE is that it can support ATM and legacy networks in the same LAN. The advantage of Classical IP over ATM is that it simplifies the protocol stack when the LAN contains only ATM switches.

5.4 Next Hop Resolution Protocol (NHRP)

The extra hop problem of the Classical model is one of the questions under consideration by the IETF *IP Over Non–Broadcast Multi–Access (NBMA) Network* (ION) working group. A new protocol, NBMA Next Hop Routing Protocol (NHRP) [74, 28, 71], is proposed which can support cut–through routing in order to eliminate these extra hops. NHRP is intended for use over both connectionless NBMA subnetworks (SMDS) and connection–oriented NBMA subnetworks (ATM), so does

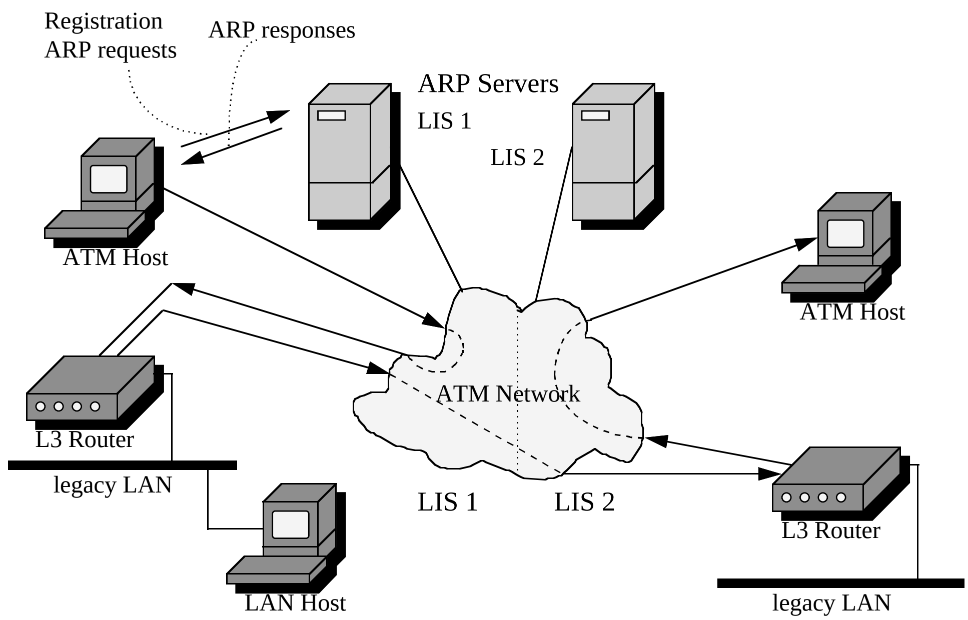


Figure 4: Inter–LIS Routing in Classical IP

not include mechanisms for connection establishment for the latter case. These must be provided by other protocols such as MPOA. Frame Relay and X.25 networks are other likely candidates for NHRP implementation.

NHRP uses Local Address Groups (LAGs) to model the NBMA networks. The main difference between the LIS model and the LAG model lies in how the local/remote forwarding decision is made. In the LIS model the decision is purely based on address information. Only nodes with the same IP network/subnet address can directly talk to each other. In the LAG model, any two nodes on the same NBMA network can establish a direct communication regardless of their IP addresses, while the local *vs.* remote forwarding decision is based upon QoS or traffic considerations. In heterogeneous networks, destinations will often lie outside the boundary of the NBMA network; NHRP has the ability to provide address resolution information for the egress router when the destination is not directly attached to the NBMA network[90].

A physical NBMA network may be partitioned into several disjoint NBMA Logical subnetworks. A NBMA Logical subnetwork is a collection of hosts and routers which share unfiltered connectivity. There are Next Hop Servers (NHSs) in the Logical NBMA subnetwork, providing NHRP service within an NBMA cloud. Each

NHS serves a set of destination hosts which may or may not be on the NBMA network. Each station on the NBMA network must have a NHS in the same Logical NBMA subnetwork which can provide authoritative address resolution information on its behalf. This NHS is the "serving NHS" of the station. Each entity which uses the the NHRP service is a Next Hop Resolution Protocol client (NHC). While NHRP can be deployed transparently in a LIS which includes ARP services and hosts which do not understand NHRP, it does require all routers on the path between the NHC and the serving NHS of the destination to be NHRP–capable.

NHCs cache the results of LIS protocol address to NBMA address resolution requests as these are learned; the information may come from NHRP Resolution Reply packets, manual configuration, etc.[72, 73]. NHRP Resolution Requests may be triggered by several different events; for instance, a host has a data packet to send, or a routing protocol update packet. When the trigger is a data packet, that packet must somehow be handled while awaiting the outcome of the Resolution Request. It may optionally be dropped, be buffered pending the Resolution Reply, or forwarded via the existing (non–shortcut) routed path toward the destination. The latter choice is recommended by the Draft, but may lead to misordering of packets once the shortcut is established. This should not be a problem for IP but may adversely affect other protocols.

In the Classical model, all data packets are forwarded hop–by–hop via intermediate routers. Routing decisions are made at each router every time a packet arrives. In the NHRP model, the same routing/forwarding mechanism is used, but not to forward the actual data packets. Instead, it is used to forward the NHRP Resolution Request packets from the source, which must be an NHC, to the serving NHS of the destination which can provide the address resolution information. The extra hop processing overhead is encountered only by the NHRP request packet. The address resolution information is used by the supported overlay protocol to establish a direct NBMA connection. Subsequent data packets are sent directly from the source to the destination via the newly established connection (Figure 5). NHRP requires a contiguous deployment of NHRP capable routers. The NHRP request packets may be dropped if not recognized by an intermediate router. In this case, packets have to be forwarded hop–by–hop just as in the Classical model.

Stable routing loops are a possibility under NHRP. This is because it violates a fundamental tenet of IP routing (as per the "Requirement for Internet Hosts" discussed above) that routing updates be sent across **all** paths through which data also flows. NHRP shortcuts are used only for data forwarding, and do not establish router adjacency. Even so, stable loops only form under relatively unlikely, even pathological network conditions. Loops are a possibility only when a "back–door" path exists between routers which is outside of, and unknown to, the NBMA network. Stable loops are likely to occur only at the boundaries between administrative domains,

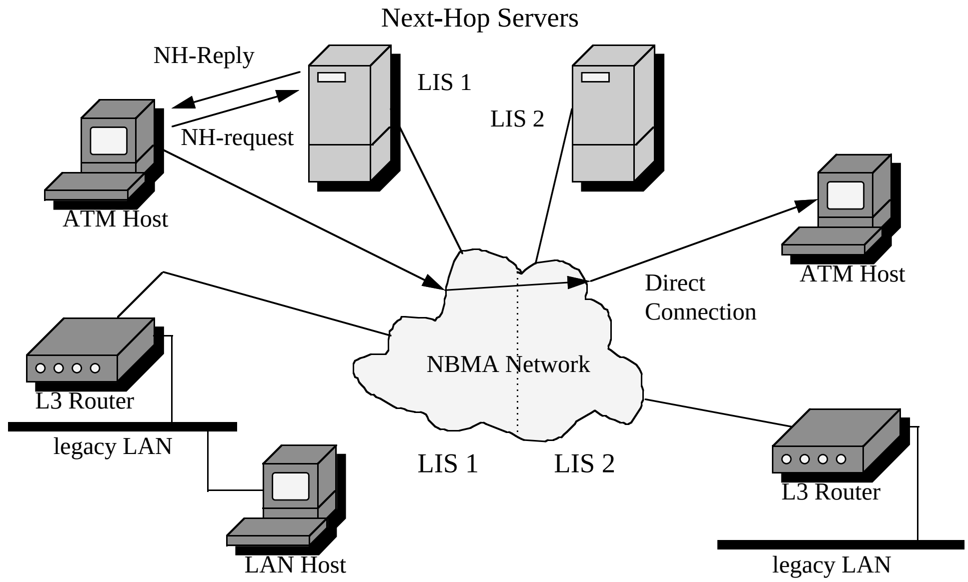


Figure 5: NHRP Solves Extra–Hop Problem

where inter–domain routing protocols lose route metrics [10]. However, the possibility is sufficiently serious that several options are under development. "Route Record" extensions for NHRP packets have been defined to aid in detecting loops, while administrative measures, particularly using routers at any boundary between administrative domains, are highly recommended.

Although NHRP solves the extra hop problem, it introduces some of its own. The first problem is the requirement of contiguous deployment of NHSs, which may not happen for a long time across the existing ATM Internet backbones. Second, the current NHRP focuses only on unicast routing. It may be possible to use NHRP to support multicast, where shortcut point–to–multipoint VCs can be used to avoid extra hops. Unfortunately, this approach is unlikely to be scalable; indeed, Ohta claims to have produced a proof that it is not [84], though his "proof" is not widely accepted among ION participants. The sender will be overwhelmed if it attempts to set up short-cuts for a very large number of receivers. This problem is even worse in the IPv6 environment where multicast is an indispensable protocol element. However, see the discussion of EARTH in Section 5.8 below.

[10]Border Gateway Protocol may have this effect

5.5 P–NNI and Integrated P–NNI

The NHRP model uses layered routing in which the internetwork layer routing and ATM routing operate independently. When a packet is to be sent across an ATM cloud, it is first routed to an ingress router of the ATM network based on IP routing. Since the IP routing protocol has no knowledge of the topology the ATM network, it may make bad decisions when chosing the ingress router. QoS routing is also a very difficult issue when the dynamics of the ATM network are unavailable. Integrated P–NNI (I–PNNI) has been proposed just to solve these problems. Before discussing I-PNNI, however, we should briefly review the ATM Forum's Private–NNI (P-NNI) on which it is based. Alles has written a very readable introduction to both P–NNI and I–PNNI [2], though it has been overtaken (at last) by the ATM Forum's creeping pace of standardization. Halpern's discussion of the architecture and status of PNNI [59] is more recent but limited in scope to P–NNI.

Private–NNI (P–NNI)

P–NNI is an hierarchical link state routing protocol and a signaling protocol, used together to establish Switched Virtual Circuits (SVCs) in a private ATM network; in this context, a "private" network is one which uses NSAP–format ATM addresses[11]. The ATM Forum's main goals in developing P–NNI are:

- Quality of Service support, and
- Universal scalability

P–NNI's signaling is an extension of UNI signaling protocols, making use of well–known VPI/VCIs to carry signaling messages. In its routing and addressing architecture, however, it draws heavily on the philosophy and world–view of the Nimrod project [31, 30]. Nimrod is an IETF–sponsored attempt to produce a "Next Generation" routing architecture, to accompany IPng. Given the slow deployment of IPv6, though, Nimrod development seems to be languishing.

Like the Nimrod work, P–NNI is a map–based routing protocol; that is, one which distributes descriptive information about the network or portions of the network, as opposed to distributing routing tables. Link State routing protocols such as OSPF and IS–IS are essentially map–based. The alternative, Distance Vector schemes (for instance, BGP and RIP) [12], can be thought of as causing routers to dis-

[11] Public ATM networks use E.164 numbering, the ITU-T B–ISUP signaling protocol and MTP Level 3 routing protocol.

[12] RIP is a distance–vector protocol; BGP is also, but it enumerates the route to the destination by keeping track of AS number sequences, alleviating some problems from which distance–vector schemes suffer.

tribute their view of the entire network rather than a map of their vicinity. Like Nimrod, P–NNI mappings abstract sections of the network which lie at differing levels of hierarchy; these hierarchical maps allow sources to select their own routes across the network. Herein lies the biggest departure from current Internet practice; paths are explicitly chosen by sources rather than fully–distributed, hop–by–hop paths in which each switch or router selects its own next hop.

Hierarchical mappings are vital to scalability, which obviously could not survive the distribution of complete network maps. P–NNI hierarchy is based on the high–order 13 bytes of the 20–byte NSAP format ATM address. This allows in principle up to 105 levels, though it isn't anticipated that more than 6 or 8 levels will ever be used. At the lowest level, logical nodes are identical to physical switches, each with a unique NSAP address[13]. Logical nodes are grouped into *peer groups*; within each peer group, all nodes use an identical topological database obtained through the exchange of full link state information. Each peer group has a logical group node known as the *peer group leader*. This is a single node which presents an abstract, summary map of its peer group to the next higher level of hierarchy. PGLs are selected through an election process. Individual peer groups are of limited size in order to avoid excessive link state information exchange (through *P–NNI Topology State Packets*, or PTSPs). By default, the peer group ID of nodes at the lowest level is given by the 12 highest order bytes of the NSAP address. This leaves one byte to specify individual switches within the peer group, limiting the number to 256. End systems obtain their network address prefix from the switch to which they are attached.

As a hierarchy of peer groups is constructed, each parent group's ID must be shorter than the prefix which is its child peer group ID. This precludes any possibility of hierarchical loops. It also allows the prefix to reflect all the ATM addresses reachable within the peer group and its children. At each level, PGLs are responsible for exchanging PTSPs with their peer nodes in the parent group in order to advertise the child group's reachability information and attributes. As well as node attributes (which include Quality of Service parameters such as available cell rate, cell delay, cell delay variation) the map includes *link* attributes, which are necessary for route selection. PGLs also distribute summary maps from the parent group and higher groups down to nodes in the child group. This allows nodes at the lowest (physical) level of the hierarchy to obtain full, though abstract and summarized, knowledge of the network.

When switches at the lowest level receive signaling requests from end systems via the UNI, they use this network information to generate *Designated Transit Lists*,

[13]The lowest level logical nodes can actually be independent networks which utilize a proprietary NNI and support P–NNI for external connectivity, rather than individual switches.

or DTLs, which are modified source routes from the end system to its desired destination. We refer to the DTL as a modified source route because, due to the summarized information used to prepare it, it cannot contain detailed information about other peer groups that a connection may have to traverse. The DTL allows a connection to reach the entry switch of a peer group. The entry switch constructs a new DTL which transits the peer group to the egress switch, then pushes this DTL onto a stack. Upon exiting a peer group, the transit DTL is popped from the stack.

DTLs are constructed using a shortest path algorithm, such as a Dijkstra calculation, over all possible paths. To do this the node requires path weights; to do this in such a way as to provide a guaranteed Quality of Service requires not simply weighting factors but a knowledge of available resources at each node on the Transit List. Each of these switches has its own *connection admission control* function, or CAC. The CAC is a local switch function, may depend upon local policy as well as physical parameters of the switch, and is in general not knowable outside the switch. However, since it is the CAC which accepts or rejects connection attempts of specified QoS, switches which generate DTLs must have some means of estimating the outcome of this function at all switches on the route. This is done through a P–NNI algorithm known as *Generic Connection Admission Control*, or GCAC. The GCAC allows any node to calculate a reasonable estimate of the CAC behaviour of another node, based on the advertised metrics of the node and the desired QoS. However, since this is only an estimate, and may be invalidated as conditions on the network change faster than PTSPs propagate, some means is necessary to salvage connections which have been rejected at a node between source and destination.

The primary means by which connections are completed in the face of CAC rejection is *crankback*. Crankback operates by the rejecting node clearing the call back as far as the node or switch which prepared the current DTL. Recalling that routing generally involves a stack of transit DTLs, this is unlikely to be back to the originating node. Given the failure information, the DTL constructor should be able to calculate another route which avoids the trouble spot. In the event that no alternate route is available, one of two actions ensues. Either the call is cranked back yet another stage, or *fallback* is invoked. Fallback is a mechanism whereby particular attributes of the requested call are relaxed, and shortest paths recalculated in order to determine whether the call can be completed with the new QoS.

The ATM Forum has chosen Q.2931 signaling for use within P–NNI, although this is actually a UNI signaling system. As such, it required a number of enhancements; for instance, a UNI has a "User" side and a "Network" side of each interface, whereas an NNI is symmetric. Additional Information Elements are also required, the Designated Transit List being the most obvious among them.

This is a very abbreviated view of a necessarily complex, end–to–end routing and signaling protocol. The full specification is published by the ATM Forum as

P–NNI V1.0 [36]. PNNI Phase 1 had not been planned to include ABR parameter negotiation, but this was encorporated by addendum [38], as was support for "soft PVCs" [37].

When it became clear that completing the Phase 1 specification was going to take a good deal longer than anticipated, the ATM Forum published a provisional, UNI–based signaling system known as P–NNI Phase 0, or *Interim Interswitch Signaling Protocol* [35] (IISP). IISP uses UNI 3.0/3.1 signaling (with switches on either side of the link arbitrarily assuming the role of "User" or "Network") and routes signaling requests on a longest–prefix–match, hop by hop basis using address prefix tables within switches. Address tables consist of entries of the form:

$\langle ATMAddress, AddressLength, InterfaceIndex\rangle$

where ATM Address is the 20–byte UNI 3.1 ATM Address, Address Length indicates the maximum number of bits to be considered in the prefix match, and Interface Index a pointer to a particular physical interface on the node. These tables contain static routes and are expected to be configured manually, or at least by means beyond the scope of the IISP specification. IISP assumes no exchange of routing information between switching systems. Clearly, IISP does not support QoS routing or crankback. CAC in switches is optional. It is a very different protocol from PNNI Phase 1 and not interoperable.

ATM Forum participants are planning the first P–NNI interoperability demonstration at Networld + Interop Tokyo in June, 1997. It is announced to include video on demand, real–time video, voice trunking, and legacy LAN interconnection using LANE, and will include ATM switches from 6 member corporations.

Integrated–PNNI (I–PNNI)

There are certainly a few "large" ATM clouds in existence (where the definition of "large" will be left deliberately vague). If and as large ATM networks become more commonplace, it will increasingly be required that routers talk to one another across these clouds. So long as IP routers have no knowledge of the internal topology of the ATM cloud, routing is at best inefficient and potentially unmanageable. At least 2 approaches based on P–NNI have been developed to resolve the dilemma. *P–NNI Augmented Routing* [25] (PAR) is the less ambitious of the two. PAR requires that ATM–connected routers run an instance of P–NNI along with their normal IP routing protocol (OSPF, RIP etc.). ATM switches in the cloud run P–NNI Phase 1. P–NNI running on the edge routers allows them to see the topology of the ATM cloud. Switches in the cloud are also aware of the edge routers, and can set up SVCs which originate and terminate at a router. Such routers are designated *restricted transit nodes*, which implies that they can never be an intermediate node in an SVC. P–NNI has been designed with the ability to carry reachability information which it doesn't understand between specific nodes, such as these edge routers; this is done using TLV encoding (type/length/value) for the IP–specific information. PAR allows routers to learn about each other across an ATM cloud without either manual configuration of PVCs or the use of an I–over–ATM protocol such as LANE, Classical IP, or NHRP, using existing IP routing protocols.

Integrated P–NNI [26] (I–PNNI) is the more ambitious alternative to PAR. I–PNNI is an extension of the P–NNI to carry internetwork layer routing information, thus allowing routing information to be exchanged between the ATM control plane and Layer 3 protocols such as IP. This results in a nearly complete integration of ATM and non–ATM networks. In this approach, ATM switches and IP routers all appear as nodes in the overall topology map, as they do in PAR. Reachability and metric information can be calculated based on this combined topology and thus can be used to find the best routes. Using I–PNNI in IP routing might make I–PNNI the first fully QoS-aware routing protocol on the Internet. It allows the router to select special paths for QoS sensitive packets. More detailed information may be found in the very useful IBM publication "Internetworking over ATM", by Dorling *et al.* [49]. Figure 6 illustrates the high–level architecture of I–PNNI.

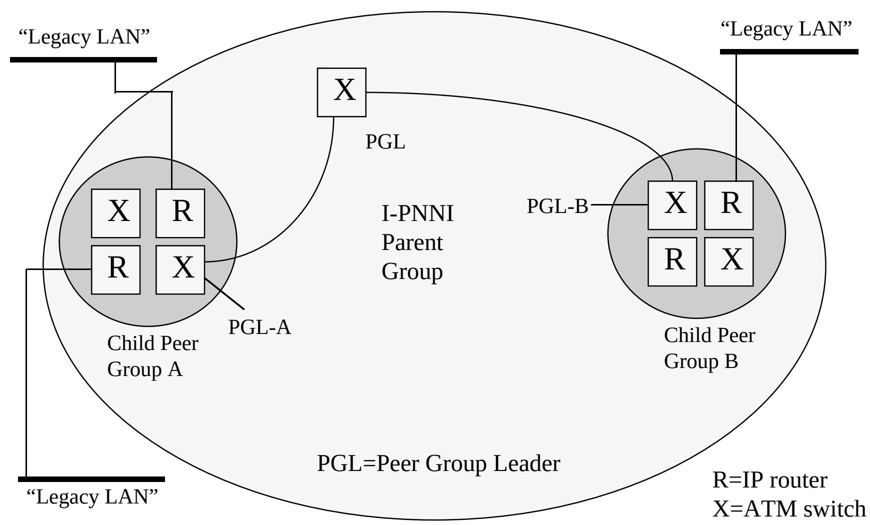


Figure 6: I–PNNI Architecture

However, deployment of I–PNNI across any substantial portion of the Internet is unlikely in the foreseeable future. It requires major changes in organization. Both routers and switches must adopt the P–NNI Peer Group hierarchy, node identifiers and peer group identifiers. Both switches and routers announce local topology via PTSPs, though I–PNNI introduces a new P–NNI Topology State Element (PTSE) which uses TLV encoding to carry IP addressing information separately from ATM addressing information. Most profoundly, standard IP routing protocols are replaced by I–PNNI routing; although I–PNNI routing does not change the way in which IP routers announce IP address reachability from the way in which it is now done under OSPF and IS–IS, it *isn't* OSPF or IS–IS, and there will be great resistance to its deployment within the Internet community. This is only likely to change if and as IPv6 becomes accepted, since as we recall IPv6's Nimrod routing and addressing is very close in spirit to P–NNI.

I–PNNI can also be used in a routing domain in a private network which contains both ATM and legacy networks. Border Gateway Protocol (BGP) can be used to exchange routing information with other routing domains. This is the most likely scenario for its early deployment.

5.6 Multi–Protocol Over ATM (MPOA) Model

Despite its name, Multi–Protocol Over ATM is an integration effort with the same underlying intention as I–PNNI – a clean internetworking of ATM networks with legacy subnetworks [27]. MPOA is a rare instance of joint development by the ATM Forum [23] and the IETF [67]. Perhaps this explains why it has progressed slowly. Again despite the name, all development to date has focussed on IP to the exclusion of other network–layer protocols. The MPOA work has several goals:

- Provide end–to–end Layer 3 connectivity across an ATM network, for hosts either directly attached to the ATM network or indirectly through routers on non–ATM IP subnets.
- Allow formation of heterogeneous IP subnets (or subnets based on other network-layer protocols) across both ATM and non–ATM networks.
- Provide direct connectivity between ATM–attached devices below Layer 3.
- Ensure interoperable, distributed routing across all network segments, using both routing and bridging information to locate edge devices nearest an addressed end system.

The design of MPOA has largely been the creation of a framework under which existing ATM elements and legacy internetworking elements can be brought together. It is a new model only in this regard. The building–blocks of MPOA have been discussed above:

- LAN Emulation (LANE)

- Next–Hop Resolution Protocol (NHRP)
- Multicast Address Resolution Server & Connection Server (MARSMCS)
- UNI 3.1 signaling (optionally, UNI 4.0) and RFC1483 encapsulation
- IEEE 802.1d spanning tree protocol for VLAN support

In this model, the behaviour of the system is modeled using Logical Components. There are two kinds of Logical Components: MPOA Servers and MPOA Clients. A collection of functions provided by a single Logical Component is called a Functional Group (FG). Forwarding and routing functions are now modeled using different Functional Groups and can be provided by different physical boxes. This allows the definition of Virtual Routers, where the route calculation is performed in a distributed fashion by a collection of route servers, which together present the behaviour of a traditional bridge/router [58]. A key benefit of MPOA is intended to be the integration of intelligent VLANs (Virtual LANs).

MPOA services are built around the concept of Internet Address Sub Groups (IASG). An IASG is formally defined as "a range of internetwork layer addresses summarized into internetwork layer routing"; it is rather like a LIS, in that it defines the logical space over which an MPOA System operates. It includes the notion of broadcast scope. It is also protocol–specific, so that when support for other than IP at the internetwork level is offered, each protocol type will define a separate IASG. Functional Groups which constitute an MPOA System are:

EDFG (Edge Device Functional Group) – internetworking connections between legacy subnetworks and ATM; constitute forwarding functions, not routing protocols.

AHFG (ATM–attached Host Functional Group) – functions of ATM–attached host participating in MPOA service.

ICFG (IASG Coordination Functional Group) – functions which coordinate distribution of a single IASG across multiple legacy LANs and ATM subnets; includes MARS functionality.

DFFG (Default Forwarder Function Group) – functions responsible for forwarding traffic in the absence of direct ATM connectivity; includes MARSMCS functionality.

RSFG (Route Server Functional Group) – functions which provide internetworking in an MPOA System; includes conventional routing protocols and provides inter–IASG address resolution.

RFFG (Remote Forwarder Functional Group) – functions forwarding traffic from one IASG to another, or between IASGs and ATM-attached MPOA clients.

Within MPOA, LANE provides the bridging function which allows connectivity within a single IASG to be distributed across multiple edge devices (ICFG and EDFG). Hosts on legacy (Ethernet and Token Ring) LANs communicate with the RSFG/RFFG functional groups as if they constituted a traditional router. NHRP provides for direct communication between ATM–attached hosts and FGs, even across different IASGs. This is known as "MPOA Target Resolution". NHRP has been extended to allow tags for data transfer packets. Tags are represented by a 4–byte field following the 8–byte LLC/SNAP encapsulation header. Tags are intended to allow the EDFG to optimize received packet processing. EDFGs are responsible for flow detection; that is, they monitor traffic to MPOA destinations to determine the number of packets per unit time. When this statistic reaches some minimum level, the EDFG must query the RSFG for the ATM address of the flow destination and set up a direct ATM VC to it, rather than continue hop–by–hop forwarding. Thus MPOA incorporates a basic element of MPLS, or more particularly IP switching, albeit through the use of the full panoply of ATM functionality.

The complexity of MPOA lies in the integration details. The basic building blocks, as we have noted, have already been described in this report. Edge devices are an area of major effort for both the standards groups and vendors; the purely ATM components already exist, and the challenges lie in the internetworking component [58]. EDFGs are in many respects logically equivalent to the control plane of IP Switches, in that they maintain flow tables and are responsible for signaling to set up flows. The ATM Forum's complex nomenclature obscures a less complicated MPOA architecture, shown in Figure 7.

Early MPOA implementations are now ready. Fore Systems will have an MPOA demonstration at Networld + Interop Tokyo in June, 1997. Their first release will include LANE, NHRP shortcuts and distributed routing. A commercial release of MPOA in ForeThought software is scheduled for 3Q97. It is clearly too early to make any estimate of the MPOA market share, but there is little doubt that it represents the ATM industry's strongest candidate for widespread internetworking.

5.7 Multi–Protocol Label Switching

Multi–Protocol Label Switching is the rather uninformative, IETF committee-generated name for a class of related technologies. A common characteristic is that the MPLS network is a transit network. End systems do not connect directly to MPLS switch–routers. Rather, the MPLS network moves data packets between ingress and egress points, where conventional routers using legacy LAN techniques de-

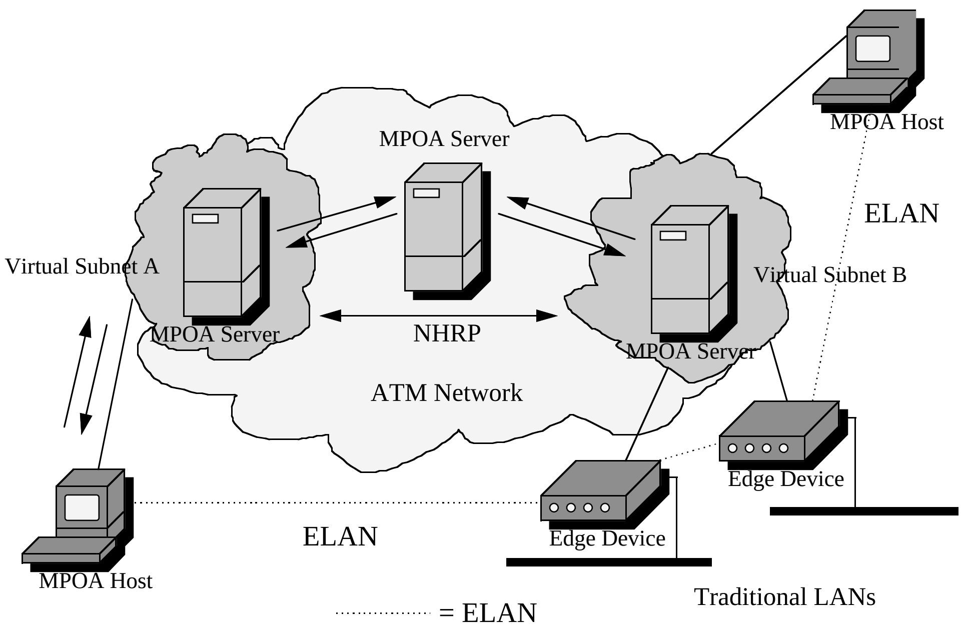


Figure 7: MPOA Architecture

liver them to the end systems. This is in contrast to such native ATM approaches as LANE and Classical IP, where ATM is used within the end system LAN itself. Two of the MPLS technologies, Cisco's Tag Switching and Ipsilon's IP Switch, originally gave rise to independent standardization efforts. However, a single Working Group was formed to deal with all of them in an attempt to ensure interoperability. The major technologies (and their primary industrial supporters) are:

- IP Switching (Ipsilon)
- CSR - Cell Switched Router (Toshiba)
- Tag Switching (Cisco Systems)
- ARIS - Aggregate Route–based IP Switching (IBM)
- SITA - Switching IP Through ATM (Telecom Finland)

These approaches have certain similarities, primarily the attempt to achieve substantial gains in packet forwarding speed by making use of short, Layer 2–like labels. However, while Tag Switching and ARIS attempt to remain independent of any lower–layer hardware implementation, IP Switching, CSR and SITA all explicitly blend an ATM (Layer 2) switching fabric with an IP (Layer 3) routing and forwarding capability. A comparison of these approaches, from the viewpoint of the CSR team, has been presented as an IETF draft [83]. The IETF MPLS Working Group has recently issued a combined Framework document [24], under the aegis of Cisco, IBM, Bay Networks and Cascade Communications, detailing the goals and requirements of Multi–Protocol Label Switching. The primary goal is to provide a core set of mechanisms which will allow forwarding of data streams [14] through the use of short, fixed–length labels associated with specific streams; the mechanisms must provide lower cost, higher performance packet forwarding than traditional techniques, while operating over the widest range of data link technologies and remaining compatible with, though independent of, any standard routing protocol. MPLS must work with both unicast and multicast streams, and scalability issues are of great concern. However, note that scalability in the Internet sense has not been designated a **MUST** issue, unlike the preceding requirements.

MPLS techniques in general are deficient in their ability to handle Quality of Service guarantees, Resource Reservation, etc., at least in early specifications and implementations. This is particularly true when flow–switching is handed off to true Layer 2 switches, as is the case for IP Switching, CSR and SITA. It is also true for ATM portions of Tag Switched and ARIS networks. Note that IP Switching has been defined only for IP, as the name implies, with other protocols such as IPX carried only by tunneling. The other technologies attempt to remain network–layer protocol independent.

Tag Switching

With its Tag Switching architecture, Cisco System, Inc., is attempting to arrive at a generic approach to one of the traditional bottlenecks of IP routers, the longest–prefix–match lookup of a packet's destination address. This architecture is intended to be applicable across all switchpoints in a heterogeneous network, whether Level 3 routers or Level 2 switches. The architecture is outlined in a Cisco Whitepaper, "Tag Switching Architecture Overview" [90]. Cisco and IBM were instrumental in founding the IETF MPLS Working Group to develop a set of open standards for this technology, and the content of the above white paper has also been published as an Internet Draft [91].

As an upper–level protocol packet enters a tag–switching capable "cloud", it is given a **tag**, a unique label which functions as an index into a Tag Information Base (TIB) residing in each Tag–Switching capable entry router. The Tag functions much

[14] A stream is defined as an aggregate of one or more flows, possibly a very large number of flows

like an ATM VPI/VCI header. It is possible for a specially designed tag–switching interior router to implement a very fast, hardware–based, L2–like switching capability for those packets which carry these tags. A software upgrade to a conventional router's operating system confers some of the benefits of the quicker lookup, albeit without the enhanced hardware switching. The system is being designed to allow the use of tags across the ATM portion of heterogeneous networks; indeed, for ATM switches the tag is likely to be mapped directly to cell VPI/VCI fields [41]. For conventional routers, the Tag is embedded as an additional protocol element, either between the Network and Transport layers, or within the Network Layer header (for instance, the Flow Label field in IPv6 [12]). The TIB associates each Tag entry with a new outgoing Tag, as well as outgoing interface and link-level information. Tags are swapped at each switchpoint, as in native ATM. Routing information resides in a Forwarding Information Base (FIB), which is constructed using standard routing protocols (e.g. OSPF, BGP). Tag-Switching capable devices exchange FIB information using Tag Distribution Protocol (TDP) [48].

Tag switches forward packets by simple label swapping, rather than by slower Network Layer forwarding as typical routers now do. The tags may be somewhat more complex than ATM VPI:VCI headers, though, and actually encompass a stack of tags. This permits information hiding (and significant simplification) when routing across enclosed domains; by using tag switches as ingress/egress routers, only the border switch–routers need maintain exterior routing information. Switches within the domain need only know about interior routing, since packets being forwarded through the domain will have exterior routing information pushed onto the tag stack at the ingress switch and popped off at the egress switch [93].

Although Tag Switching aims to lend Layer 2–like label swapping to IP routers, the developers have been less successful in mapping this technology directly onto ATM switches. This must be considered an absolute requirement, since the intention at the outset was not only to gain forwarding speed in routers but to make transparent a heterogeneous network of routers and switches. "Use of Tag Switching with ATM" [41], another IETF draft, defines the manner in which tags are carried in VCI and possibly VPI fields of ATM cells. Superficially this is a trivial mapping, since the role of tags and VPI/VCI fields is nearly identical, and their method of label–swapping at switch interfaces is conceptually the same though differing in detail. However, there are two areas in which operation of normal ATM switches does not follow the requirements of Tag Switching: flow merging and TTL decrementation. Tag Switching is thought of as destination–based. Multiple flows which converge to a single destination may be merged into a single stream at various merge–points along a multipoint–to–point tree. Normal ATM switches are not capable of VC–merging, however, since lacking demultiplexing information within cells, it's not possible to separate cells from different frames or packets should they become merged into a single Virtual Circuit. Tag Switching depends for scalability on the ability of a node to reach N other nodes via $O(N)$ switched paths, as is the case for IP routers. The requirement to map a single Tag Switched stream into multiple ATM Virtual Circuits leads to $O(N^2)$ paths for a full–mesh connection, which is hardly scalable to Internet levels.

TTL decrementation is the normal method used in routers to limit damage due to routing loops, but is not applicable to normal ATM switches. Native ATM routing procedures depend upon preventing loops (Section 5.5) while IP routing accepts transient loops. When Tag Switch–capable routers act as edge–routers for a cloud of ATM switches, they may handle this problem by pre–decrementing the TTL field at the IP level by the number of hops across the cloud, pushing the exterior routing tag on a tag stack, also at the IP level, then attaching the interior routing tag VPI/VCI before injecting the frame into the ATM cloud. Whether there is a need to prevent loop formation in an MPLS network, whether it is sufficient to detect and repair loops quickly, or whether the ability of normal IP routing protocols to eventually detect loops can be augmented by techniques which minimize damage to network throughput and overall performance, are the subjects of much contention in the MPLS Work Group. Discussion of these issues forms a large part of the Framework document.

The Framework document requires that "MPLS MUST be compatible with the IETF Integrated Services Model, including RSVP" [24], and this is supported in Tag Switching. The beginning of this support may be seen in its support for Classes of Service [70]. This is reminiscent of Cisco's COS (Class of Service) capability in its NetFlow IOS router software, but is far from the capabilities required to implement RSVP, for instance.

Multicast routing is supported by the ability to associate tags with multicast trees. The informational draft indicates that tag switching can support a diverse routing functionality. A tag might be associated with a group of routes, functioning thus much like an ATM Virtual Path. It might also be used in destination-based routing (forwarding based on the destination address in the packet as well as information in the FIB), in the manner of existing IP routers. Unfortunately, the informational draft is deficient in detail as to how these diverse functionalities will precisely be accomplished. For instance, it is claimed that tag switching on an ATM switch may require it to maintain several tags associated with a single route or group of routes with the same next hop, in order to avoid interleaving of cells from different packets. It is not clear from the discussion how even a group of tags will allow ATM's connection-based nature to be accommodated to a route-based paradigm. However, this lack of precision is likely for proprietary reasons rather than indicating a failure of reasoning in the protocol.

Cisco holds a dominant position in backbone routers as well as in campus-level routers, and now has StrataCom's BPX line of ATM switches as well. The company has announced that the Tag Switching architecture will be incorporated in upcoming versions of their IOS technology, so there are no issues of protocol formalization involved - this protocol WILL be used on the Internet.

Aggregate Route–Based IP Switching (ARIS)

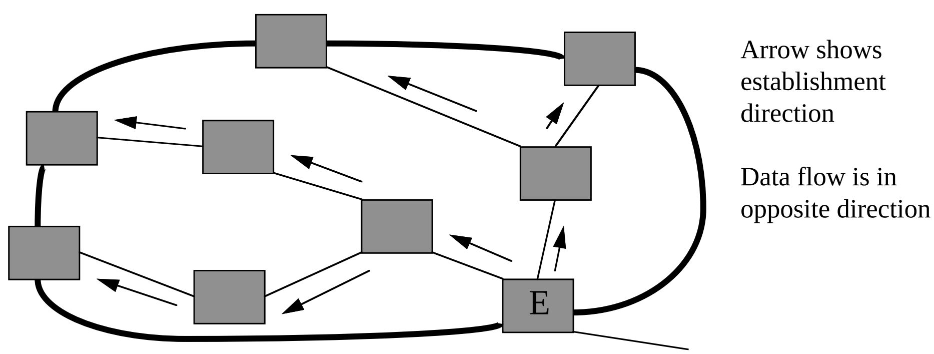
ARIS [118, 114] is almost solely sponsored by IBM, though it too is under development as an open IETF standard [54]. It is intended for use with switched network technologies, whether ATM, Frame switches, or LAN switches, and permits L2 switching to be used for IP datagram forwarding. Unique among MPLS proposals, it greatly encourages the use of ATM switches which have been specifically designed to accomodate VC–merging (though this is not an absolute requirement). That is, an ARIS *Integrated Switch–Router* (ISR) should have sufficient buffering to permit assembly of AAL5 PDUs, and a control plane which understands packet–level queuing and all the related queue–control algorithms. Packets arriving with different VP/VCs can be forwarded onto a single VP/VC (merged) by being retransmitted sequentially once an entire datagram has been received, without any cell interleaving. An ISR does not necessarily have to be capable of ATM Segmentation and Reassembly, however, as cells may remain buffered independently in a special queue rather than actually being reassembled into a PDU. In the absence of VC–merge capability the ability to merge switched Virtual Circuits must be implemented through the use of Virtual Paths (VP); this alternative scales poorly, however.

ARIS differs from Tag Switching in its use of a Route–Based paradigm rather than a Flow–Based one. This is quite a significant difference. A route in this sense is rather like a multicast distribution tree, rooted at the egress point, and traversed in reverse. The egress point is specified by an egress identifier, which may be one of:

1. IPv4 destination prefix
2. egress router IP address
3. OSPF Router ID
4. multicast (source, group) pair (DVMRP, MOSPF, PIM–SM, PIM–DM)
5. multicast (ingress–of–source, group) pair (MOSPF, PIM–SM)
6. as yet undefined, but likely to include IS–IS NSAP addresses, IPv6 destination prefixes, others.

Using the IP destination prefix, packets from any ingress point forming a leaf on the route tree, and intended for that egress point's destination prefix, are switched and merged to the root egress ISR. This is useful in a relatively small (campus or small enterprise) environment running RIP, but does not scale to backbones. The egress IP address is used primarily for BGP protocol updates. OSPF Router ID is the most generally useful, supporting both IPv4 and IPv6, and allowing aggregation of of all OSPF–routed datagrams.

An ARIS network (or network comprised of ARIS capable ISRs) establishes switched paths to "well–known" egress points, independently of any traffic. These egress points are established through the operation of standard Layer 3 routing protocols (OSPF, BGP, etc.). It is actually the responsibility of the egress ISR to initiate the path setup by sending messages (*Establish* messages) to upstream neighbors (see Figure 8). These neighbors forward Establish messages upstream in Reverse Path Multicast style, so that eventually all ARIS ISRs have switched paths to every egress ISR.



* Switched path established to each egress node (E)
* Switched paths follow IP forwarding path
* Single path for all destinations behind common egress
* One tree rooted at egress

Figure 8: IBM ARIS Switched Paths

An important element of ARIS is that switched paths are guaranteed to be loop–free, despite use of standard IP routing protocols. Each ISR appends its own "ISR ID" to Establish messages it forward, in a manner similar to IP's "Route Record" or BGP's AS_PATH attribute. It can then determine whether the Establish message has passed before, hence is looping, and refuse to continue the path. The loop will eventually be deleted as the Layer 3 routing converges, whereupon the path can be properly established. When IP routing initiates a path change, the ISR local to the change is required to "unsplice" the obsolete path and create a new one with a fresh set of Establish messages. Ingress ISRs are assumed to have true IP forwarding capability, and are required to decrement IP datagram TTL fields by the number of switched hops along the path, plus 1. This gives compatibility with non–ARIS, IP networks which use TTL for limiting distribution scope as well as loop control.

ARIS supports source–routing and multicast (Point To MultiPoint). P2MP trees are initiated at the ingress ISR, but each branch of the P2MP tree must become part of a switch path tree rooted at the egress ISR. It is unclear whether the scalability properties of this interlinked tree complex have been fully investigated.

ARIS path information is soft state, maintained only as long as ARIS messages are seen within a timeout period; *KeepAlive* messages are used to maintain state in the absence of real traffic.

Ipsilon IP Switch

IP switching is a generic term applicable to many of the MPLS technologies, but as a name, "IP Switch" is a particular product group from Ipsilon Networks. Despite this, Ipsilon Networks is releasing specifications and protocols as IETF RFCs. Newman *et al.* have published a decent overview of the Ipsilon technology [81]. The IP Switch is composed of an ATM switch and an IP Switch Controller. The IP Switch Controller is a routing and forwarding engine linked to the switch via one of its OC-3c ATM ports. Although the IP Switch is more firmly tied to ATM technology than, for instance, ARIS or Tag Switching, the ATM switch itself is used only as a switching fabric. All normal ATM signaling and control plane functionality are abrogated.

Like ARIS ISRs, IP Switches interconnect IP subnets, not end systems. The FAS200 [15] connects one or two OC-3c ATM links with up to four 10/100 Mbps Ethernet LANs. It functions as an ingress or egress switch/router. The ATM1600 IP Switch [16], with 16 OC-3c ATM ports, is purely an interior switch/router. IP Switches depend upon flow classification, similar to Tag Switches. Flow classification is described as a local policy decision, but in general will depend upon some combination of source IP address and port number, destination IP address and port number, and protocol type. Datagrams which are part of flows which are expected to persist for a significant time will be forwarded differently from datagrams which

[15] *IP Switch FAS200*, ©Ipsilon Networks Inc., `<http://www.ipsilon.com/products/fas200.htm>`

[16] *IP Switch ATM1600*, ©Ipsilon Networks Inc., `<http://www.ipsilon.com/products/atm1600.htm>`

are purely connectionless and non–flow–associated.

Traffic likely to qualify as a persistent flow is:

- FTP transfer
- HTTP
- Telnet
- Multimedia audio/video

while non–flow or short–lived flows will comprise traffic such as:

- DNS lookups
- SMTP transactions
- SNMP transactions
- NTP

By startup configuration, an IP Switch network forms a full–mesh of Virtual Channels (over well–known VPI/VCIs) between all switches; these VCs logically connect IP layer controllers on all switches together through the ATM switch fabrics. Switch Controllers use Ipsilon's Generic Switch Management Protocol [80] (GSMP) to communicate with and control the ATM switching fabric over their OC-3c connection. Upon startup, no flows are yet known, the IP Switch network looks like a connectionless net of IP routers, and runs standard IP routing protocols. IP traffic entering the IP Switched network via one of its ingress routers is forwarded hop–by–hop to its desired egress, via the VC web between Switches.

IP Switches use the Ipsilon Flow Management Protocol [78] (IFMP) to identify traffic which is part of a relatively long–lived "flow" and request that the upstream node set up a dedicated virtual circuit for it. The ATM switch is instructed to switch it to the IP forwarding engine via this new VC rather than the common, connectionless VC used for incoming default IP traffic. However, datagrams from this flow are still forwarded over the default outgoing IP Virtual Channel to the downstream node. Figure 9 indicates the control and data flows, and protocol applicability, for a schematic IP Switch.

Now when the downstream node has also identified this as a persistent flow, and requested (via IFMP) that a unique VC be set up to carry it, the IP Controller recognizes the opportunity for cut–through. This flow enters the ATM switch on a unique VC and is switched up to the IP Controller on this VC; forwarded normally

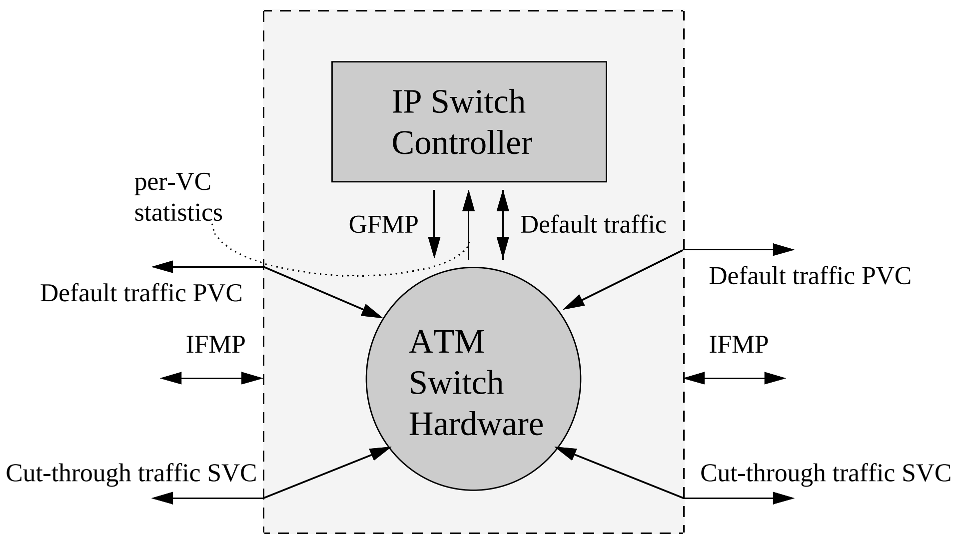


Figure 9: IP Switch Architecture

there via another unique outgoing VC back to the ATM switch, and then out to the downstream node. The IP Controller instructs the switch to logically connect the incoming VC with the outgoing VC within the switch fabric, without being routed up to the IP Controller, and the cut–through route is established [79]. At this point, the IP Controller no longer has to perform L3 forwarding for any datagrams from the flow; indeed, it no longer sees this flow at all. However, ATM–level statistics for the associated VC are available to the IP Controller via GSMP. Flows are soft state, maintained while there is traffic by periodic redirect messages and deleted in the absence of traffic by reclaim/reclaim ack exchanges. IFMP incorporates an adjacency protocol to allow switches to detect any changes in their neighbors and clear any flows which involve an altered link. Figure 10 indicates the overall architecture of a network composed of IP Switches.

An additional benefit is conferred by a change in encapsulation. IP datagrams on the default VC are RFC1483 [60] LLC/SNAP encapsulated over the AAL5 link, while the cut–through VC uses "per–VC" encapsulation; all IP header fields are removed from datagrams by an IP Switch which accepts a redirection request and restored by a switch which issues a redirection request. There is a small gain in bandwidth efficiency, but a more important gain is in security; it becomes impossible for a malicious user to establish a flow and then change IP headers in some way which does not invalidate the flow but would still allow unauthorized access to downstream resources. It is also the responsibility of these switches to adjust the

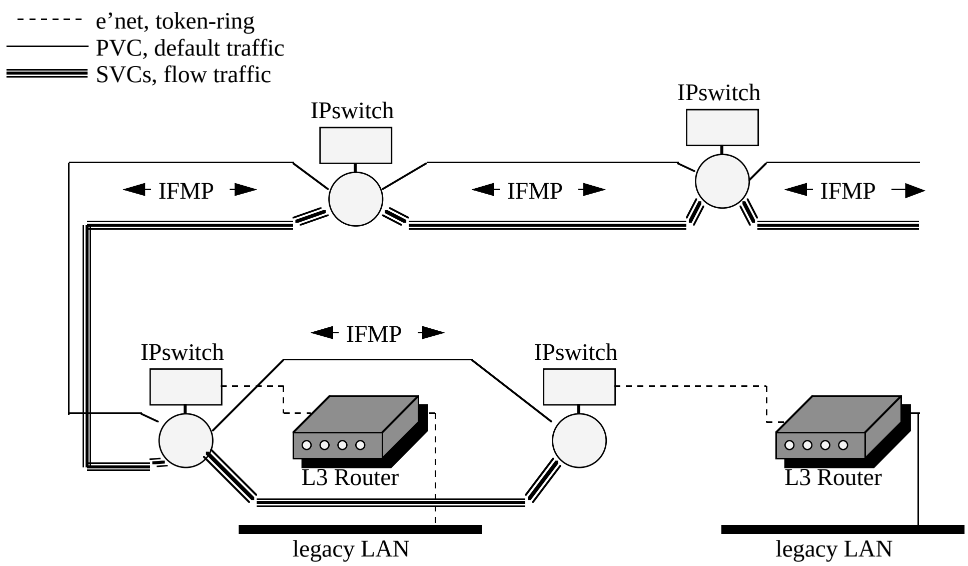


Figure 10: IP Switch Network Architecture

TTL to reflect the number of switched hops between them.

Although Ipsilon uses the term "cut-through", it should be noted that the ATM data path between routers in the IP Switch network is exactly the same as the hop–by–hop IP path. Cut–through is more commonly used to refer to a situation where ATM–specific routing protocols (such as NHRP) allow end–systems to request direct connections between them without passing through intervening IP routers.

The major advantages of the IP switching approach are first, that for traffic which recognized as constituting a flow, the overhead of higher–level packet forwarding, primarily address lookup via longest–prefix matching, is avoided; and second, that the delay required in order to reassemble and then resegment each packet in the IP router is eliminated. Both reduce latency through the switch/router, while the latter may also reduce delay variation. Standard IP routing protocols apply without modification, and ATM's end–to–end call setup is avoided. However, since the routing aspect of this protocol functions as an overlay, it takes no advantage of ATM's ability to offer QoS guarantees; it depends upon generic RSVP and local "flowspec" implementation policy for this. At present, it is unclear how this approach will scale; it seems to require one virtual circuit per flow. Depending upon flow duration, and the additional flow state holding time imposed by the redirect/reclaim timescales, it is possible that this will lead to "VC starvation" in large clouds with many edge routers. Simulations based on traces from the MAE-West FDDI ring indicate that something like 80% of packets, representing 90% of all bytes, would be qualified as flows; critics argue that this is a poor representation of internet backbone traffic and the actual fraction of traffic in flows will be much lower, reducing the benefit from this approach. IP switching is an open protocol and may eventually be supported by multiple vendors. However, IP switches can only set up flows in concert with others of their kind, and can interoperate with standard ATM switches only over manually–configured PVCs, so their use is limited to relatively small private domains unless and until the approach becomes widely disseminated.

Cell Switching Routers (CSR)

The Cell Switch Router (CSR) architecture is in large measure the work of Ohta at the Tokyo Institute of Technology, but has been further developed and refined to a semi–commercial product by Esaki, Katsube and others from Toshiba Corporation. It is very similar in many respects to the Ipsilon IP Switch. It differs in that its intention as described in the Toshiba whitepaper [50] is to connect ATM Local IP Subnets (LIS), running LANE or RFC1577 Classical IP over ATM, **not** IP subnets with non–ATM datalinks using standard IP routers. Connection between the ATM LIS and non–ATM networks are via standard routers, as per LANE or RFC1577 specifications. Signaling between CSRs uses the UNI 3.1 Q.2931 standard, and address resolution depends upon ATMARP and InATMARP servers. Ohta is a stern expositor of the virtues of the CATENET model, as described in RFC791 [88], the DARPA IP specification. That is, the network and datalink layers are topographically identical within an IP subnet, and IP subnets are connected only by routers. Hence this is the model used for the CSR. It does not permit decoupling of the L2 and L3 topologies, which is quite common in non–routed subnets, and specifically deprecates NHRP–type shortcuts.

Like the IP Switch, the CSR is capable of doing both cell switching and IP forwarding. It is unnecessary to repeat the description of the default (IP forwarding) and ATM–level connection setup procedures, as they are very similar to what we have described above for the IP Switch. The conditions under which specific ATM VCs are set up have not been clearly defined in the CSR literature. RSVP RESV packets are one clear form of “trigger packet” which is intended to result in a connection, and the notion of flows, whether based on traffic analysis or IPv6 flowspec is supported but not specified. A *Flow Attribute Notification Protocol* [76] (FANP) has been published as an IETF RFC; it details the format used between CSRs to set up flows by Q.2931 signaling.

The advantages and disadvantages of the CSR are similar to those of the IP Switch. Because the CSR connects LISs which use RFC1483 LLC/SNAP encapsulation [60], it is possible to route non–IP network protocols, which the IP Switch cannot do. See Figures 11 and 12, due to Noritoshi Demizu of the Sony Corp., for a comparison of CSRs with the other MPLS technologies. These tables indicate FANP runs over IP, while a perusal of the RFC indicates otherwise, and also make no mention of the RSVP–based QoS support in CSR mentioned above.

The CSR development group has been very active in devising methods and specifications to broaden the application of MPLS techniques specifically over ATM, for instance "Router Extensions for ATM" [64] and a series of publications relating to signaling for MPLS over SVC: "IP Address Resolution and ATM Signaling for MPLS over ATM SVC Services" [51]; and "Another ATM Signaling Protocol for IP (IP-SVC)" [55]. However, the Toshiba Corp. CSR has had little or no commercial penetration, at least in North America, and its influence is primarily intellectual.

Other Models: SITA

Switching IP Through ATM (SITA) is a proposal for connecting together a collection of edge routers across an ATM network. Its originator is Juha Heinanen of Finnish Telecom, an IETF "eminent personality". It was done in response to discussion in the MPLS Working Group, in order to illustrate a simple solution to the problem of VC merging on standard, unmodified ATM switches. When more than one Virtual Circuit arriving at a particular switch/router requires to be sent to a single destination, SITA proposes using a unique Virtual Path ID (VPI) for that route alone, and multiplexing the VCIs inside it. Considering that the VPI space in ATM cells at the UNI is only 8 bits, this is not exactly a scalable model, and indeed it was probably never intended for implementation. Heinanen's proposal may be found on the WWW at `<http://www.cs.ubc.ca/World/mjmccut/sita.html>`.

5.8 Multicast Over ATM

Multicast is a field under very active development within the IP community. While it is sometimes invoked as a critical test of ATM's suitability as a medium to carry IP, its use in the Internet as a whole is still largely experimental. No universal solution yet exists. There are two very different regimes in which different protocol sets are being developed: Dense Mode and Sparse Mode [101]. The Sparse Mode architecture [45] and protocols [52, 44] as well as the Dense Mode protocols [46] are defined in articles and Internet Drafts, not yet as Standards (RFCs).

Multicast as it is evolving in the Internet is something of a hard problem for ATM. While all current ATM switches support some form of Point–To–Multipoint forwarding, like all ATM connections this is intrinsically unidirectional. Moreover, with standard ATM switches it is impossible to directly support Multipoint–To–Point functionality when using PDUs of greater than single–cell size (e.g. AAL5)

Multi Layer Routing

Comparison Table

Table courtesy of Noritoshi Demizu (SonyCSL)
<http://www.csl.sony.co.jp/person/demizu/inet/mlr.html>

	IP Switch (Ipsilon)	Cell Switch Router (Toshiba)	Tag Switch (Cisco)	ARIS (IBM)	SITA (Telecom Finland)
Datalink Layer	ATM (w/o VPI/VCI)	ATM, FR, etc. (CO datalinks)	ATM, FR, ethernet, etc (CL & CO datalinks)	ATM, FR, etc. (CO datalinks)	ATM (or switches with two levels of tags)
Network Layer	IPv4, IPv6	IPv4, etc	IPv4, XNS, apple, etc (Protocols in Cisco IOS)	IPv4, IPv6, etc	IP, etc.
Between L2 & L3	(none)	(none)	a small "shim" tag header for label-swapping	(none)	(none)
Hierarchical Switching	(none)	(none)	with a stack of tags	(none)	(none)
Cell-VC Merging	(no need)	(no need)	(requesting?) multiple VCs instead of merging, or L3 routing?	assumed (alt. VP merging or end-to-end VCs)	(no need)
ATM Signaling	no but DEC has it	yes	yes?	yes?	?
VC Setup Protocol					
Protocol Name	IFMP and IFMP-C (over IP)	FANP (over IP)	TDP (over CO-transport)	ARIS protocol (over IP)	(no setup protocol)
State	Soft-State	Soft-State	Hard-State	Soft-State (&Hard-state)	-
Assignment	by Downstream	by both	by both?	by Downstream (mainly)	-

Figure 11: Multi Layer Routing Comparison Table A.: General and VC Setup

VC Setup Protocol					
Protocol Name	IFMP and IFMP-C (over IP)	FANP (over IP)	TDP (over CO-transport)	ARIS protocol (over IP)	(no setup protocol)
Best Effort Traffic	**Traffic-Based** per Detected Flow - addr, proto, port - packet counting expired after 60s	**Traffic-Based** per Detected Flow - addr, proto, port expired after 120s	**Topology-Based** per Destination Prefix - routing table entry using routing hierarchy	**Topology-Based** per egress ID - each egress has a tree - reversed path multicast tree algorithm	**Topology-Based** VP switching - VPIs for all egresses - VCIs for all ingresses
Percentage of L2 switching (B.E.)	80%-90% ? (long-life pkt only)	80%-90% ? (long-life pkt only)	Any data? (all except "hop-by-hop" packets and route aggregation)	Any data? (all except "hop-by-hop" packets and at border routers)	Any data? (all except "hop-by-hop" packets and at border routers)
Loop			Detected by - hop count	Prevented by - initiated only by egress - ISR ID path	
QoS Support	**Reservation-Based** per Reserved flow? - QoS by priority?		**Reservation-Based** per Reserved flow? (using routing hierarchy)		multiple VPIs with different QoS classes for each egresses
Multicast					
Other use			Traffic Tuning		

Termninology					
	IP Switch	Cell Switch	Tag Switch	IP Switch?	
		IP level routing	Layer 3 routing		
		ATM level routing	Layer 2 switching		
	cut-through	bypass			
	association		binding		
Routers	IP Switch + IP Switch Gateway	Cell Switch Router + Edge Device	Tag Switch + Tag Edge Router	Integrated Switch Router	

- IFMP = Ipsilon Flow Management Protocol
  IFMP-C = IFMP Client
- FANP = Flow Attribute Notification Protocol
- TDP = Tag Distribution Protocol

Figure 12: Multi Layer Routing Comparison Table B.: Mapping Method

due to the so–called "VC Merge" problem. That is, when PDUs are larger than a single cell, the destination host depends upon receiving all cells of a given PDU in order and unmixed with cells from other PDUs on a given Virtual Circuit. This is because AAL5 lacks information on a per–cell basis for demultiplexing different PDUs within a single VC. Thus in order to support Multipoint–To–Point, or more generally Multipoint–To–Multipoint, ATM generally needs to utilize multiple Virtual Circuits. This leads to scaling problems and the possibility of VC exhaustion. Another related incompatibility has been that IP multicast is receiver–initiated while, until UNI 4.0, ATM signaling supported only sender–controlled group membership. Hence alternative approaches have been developed. RFC1754 (Recommendations for the ATM Forum's Multiprotocol BOF) [67] was a first attempt to reconcile the IETF's ATM multiprotocol developments with those of the ATM Forum, and perhaps the first IETF standard to deal with multicast in an IP/ATM environment. In "Multicast and Multiprotocol Support for ATM based Internets" [6], Grenville Armitage has reviewed the multicast work being done in the IETF ION working group.

There are at present two ways to implement multicast service over ATM – ATM Multicast Servers (MCS) and ATM VC meshes. In the former, multicast packets are first sent to the server, and then are redistributed to all the receivers. In the latter, *each* sender sets up a Point–To–Multipoint VC to all receivers. The Multicast Address Resolution Server (MARS) protocol [7] details the Address Resolution aspect, proposed for use in a Classical IP over ATM environment. It can support either the VC mesh model or the MCS model. There is at least one MARS server in each Cluster (usually the same as a LIS) which maintains a list of all the local receivers in each the groups. When nodes join or leave a certain multicast group, MARS_JOIN or MARS_LEAVE messages are sent to the MARS server and are further forwarded to all members of the group. Farinacci, Meyer and Rekhter [53] have proposed using PIM–SM's "explicit join" mechanism to support efficient intra–LIS multicast over P2MP VCs without the need for MARS.

The VC mesh mechanism is suitable for small groups, and is consistent with the Cluster or LIS, but will not scale to large cloud or Internet proportions. In the VC mesh model, when a host wants to send to a group, it sets up its own Point–to–Multipoint (P2MP) VC for each group it is sending to (conventional IP multicast routing protocols could be used to forward multicast traffic between Clusters).

The MultiCast Server (MCS) model [110] extends the MARS model to use Servers rather than VC meshes. The MCS establishes a P2MP tree with itself as the source, to all registered multicast group members in the LIS. There may be more than one MCS within a LIS for fault tolerance, but only one is active at any given time. This alleviates the requirement for full–mesh connectivity between all members of the multicast group, which may be an inefficient use of ATM resources. This arrangement is illustrated in Figure 13.

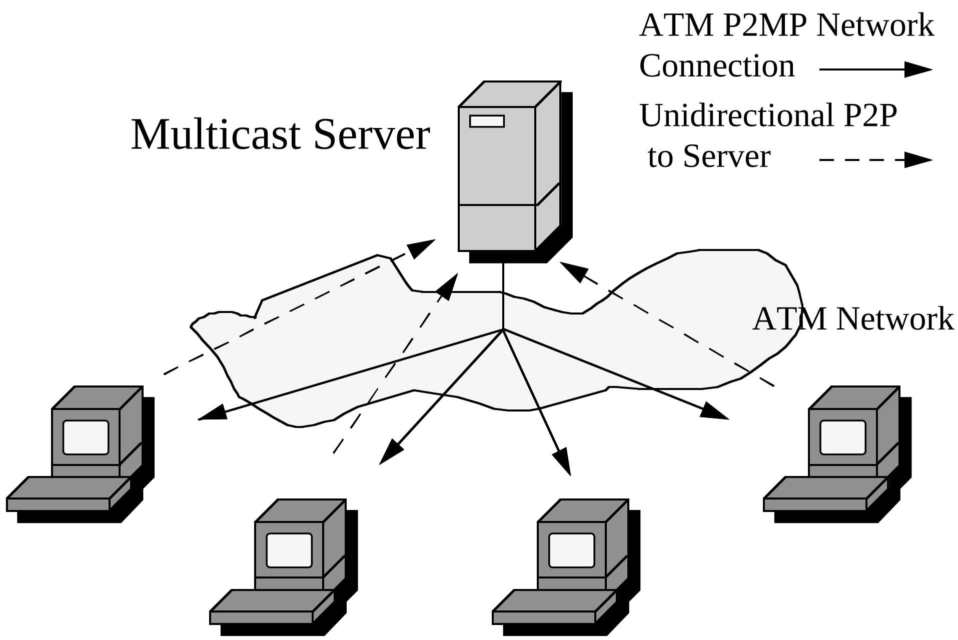


Figure 13: MARS Multicast Server Operation

Dense multicast groups over large ATM clouds are still not supported. However it is possible to use NHRP to support sparse–mode PIM on an ATM network since all the messages involved in PIM are unicast. Armitage's VENUS (Very Extensive Non–Unicast Service) model [4] is a "strawman" proposal intended to show the difficulty of extending NHRP shortcuts to MARS multicast. Smirnov's EARTH (EAsy IP multicast Routing THrough ATM clouds) [104] is a response to VENUS. It describes a means of implementing shortcuts which is based on Multicast Logical IP Subnets (MLIS). MLISs span the entire physical ATM network, even though that network may be partitioned into logically disjoint unicast LISs. This proposal uses EARTH servers which provide IP Class D address resolution to ATM addresses.

5.9 IP Integrated Services over ATM

The concept of IP Integrated Services has largely been dealt through Resource Reservation, and the Resource ReSeRvation Protocol (RSVP) [119]. RSVP fits reasonably well with the design philosophies of the Internet; it maintains only soft, periodically refreshed state information in the intermediate nodes, while most of the control states and related complexities are maintained in end systems. This allows RSVP to adapt automatically to route and membership changes. However, it is an unfinished specification, has not been widely implemented, and it is not clear whether and how it will function in a global Internet, for administrative as well as technical reasons. As previously mentioned, there is a substantial sentiment within the IETF that "best effort" is good enough so long as "infinite bandwidth" is available. But of course there is no agreement on this approach, and RSVP is the alternative.

It took the RSVP working group more than 2 years to finalize the protocol. Now the final version (version 13) of the RSVP draft standard [21] is available and is believed to be on track for RFC standardization soon. Alpha versions of RSVP implementations are also available.

RSVP is not a routing protocol. It may operate together with any routing protocols available on the internet. Its aim is to accept a path (or distribution tree), however obtained, and set up and maintain resources over that path for the use of the entity which requested that reservation. It also introduces the concept of *filtered reservations*. It is possible to make reservations which may only be used by packets from certain specified sources, and the list of allowable sources may either be fixed or dynamically changeable with time. Such flexibility is particularly important for multicast groups. A fixed filter allows switches or routers to merge individual reservation requests, knowing that the characteristics of the reservation will not change. Conversely, a dynamic filter reservation gives the receiver the ability to change sources from time to time, or "change channels".

RSVP features receiver–initiated reservation which means that the receiver is responsible for joining the distribution tree and setting up reservations on all the intermediate nodes. This decision is consistent with the receiver–initiated establishment of multicast distribution trees. The decision also enables RSVP to accommodate heterogeneous receivers on the same distribution tree, which leads to an extremely flexible and scalable design.

At first glance, implementing IP Integrated Services over ATM seems a natural match, since ATM's built–in support for QoS, and native connection–orientation, were designed with service guarantees in the forefront. On deeper consideration, though, this seems to present intractable problems, given the different logistics of the two technologies. Both the underlying network (ATM) and the overlay network (IP) have mechanisms for implementing service guarantees. Since these mechanisms differ, there is no agreed–upon "division of labor" between the layer mechanisms or means of communicating QoS parameters, and most importantly, no agreement that Layer 3 entities *should* be able to communicate QoS parameters to one special type of data link (Layer 2) network, there are clearly many hurdles to overcome. Still, a good deal of the work has already been done (see discussion of the

ISSLL contributions below).

Many issues related to running RSVP over ATM are still unsettled. In an IP-over-ATM environment, ATM SVCs are usually used to support QoS guarantees and RSVP is used as the internet level signaling protocol to convey the QoS requirements. Several issues have to be addressed in such an environment. The first issue is when and where to use SVCs.

In the LIS model, this is not a problem since full-mesh VC connection within the LIS is assumed, and hosts never talk to anyone outside the LIS except through IP routers. RSVP over ATM will operate in just the same way as over legacy networks. In the NHRP model, SVCs are used to establish direct short–cut links between sender and receiver when the data volume is high or the QoS requirement is tight. However only unicast short–cut VCs may be used since RSVP needs bi–directional connections to transmit RESV and PATH messages, while ATM point–to–multipoint VCs are unidirectional[17]. Since NHRP for multicast is still not a reality, this is probably a good thing. Heterogeneity is another problem in RSVP over ATM. RSVP allows different receivers of the same session to have different QoS requirements, or even have different QoS classes. The ATM UNI 3.x and 4.0 only support homogeneous QoS for all receivers. To solve this problem, a VC which has the maximum QoS requirement of all the receivers has to be established. Of course, some resources are wasted, and some intended "best effort" receivers may get real–time links, resulting in an actual degradation of the desired QoS. So in the presence of best–effort receivers in a session, it may be desirable to set up "best–effort" VCs to each of them.

RSVP also addresses QoS renegotiations and dynamic membership, currently not supported by ATM networks. QoS renegotiation and membership changes have to be implemented by setting up a new connection and tearing down the old one.

These issues and questions are being dealt with in ongoing work in the IETF Integrated Services over Specific Lower Layers (ISSLL). This work is directed toward mapping RSVP to ATM UNI services. Borden *et al.* [17] outline the issues raised by RSVP–over–ATM; Berger's draft [14] gives guidelines for implementing RSVP over ATM. Berson and Berger [15] and Williams *et al.* [116] provide methods for using ATM VCs with QoS under RSVP. Borden and Garrett provide suggestions for service mappings between IP Integrated Services and ATM Quality of Service [56]. Crawley [40] narrows the focus to IP Integrated Services over LANE.

[17] In the presence of short–cut routes, the up–stream and down–stream route may be different. Thus RESV messages might arrive at the wrong router or wrong interface. The misrouted RESV packet is actually not a problem since RSVP will forward the RESV to the correct destination, and the RESV message will not affect the out-going interface of the previous–hop node.

Real-Time Protocol (RTP)

Real–time traffic is often considered as a prime example of the type of service which requires Quality of Service guarantees to function well. RTP [99] is a IP–based transport protocol suitable for transmitting real–time data such as video and audio streams. It is primarily designed to satisfy the needs of multi–participant multimedia conferences. RTP by itself can not guarantee the timely delivery of data packets. Instead, it depends on lower level protocols such as RSVP to fulfill the QoS requirements and actually provides a mechanism to deliver the QoS requirements from the application to the underlying network services. An auxiliary control protocol (RTCP) [97, 98] is also defined to monitor the data delivery. RTP and RTCP packets are usually transmitted using UDP service.

RTP includes several fields in its packet headers to specify the time and synchronization attributes of the data. These include a payload type field which indicates the format of the payload, a sequence number field used to indicate the location of the packet in the stream, a timestamp field which reflects the sampling instant of the first octet in the data packet and a synchronization source field to identify sources that should be synchronized during playback. The default payload types are defined in RFC1890 [98]. Specifications have been developed for encapsulating MPEG1/MPEG2 audio and video [63], JPEG video [13], Sun CellB video [106], H.261 video streams [112], and, more generally, layered multimedia [107]. Two different levels of MPEG encapsulation have been developed. The lower level is suitable for implementation on existing IP connected workstations, and defines encapsulation of compressed audio and video data in the form of MPEG "Elementary Streams". The more complete specification encapsulates both MPEG Transport and Program Streams, and makes available the full semantics of the MPEG system.

Examples of RTP encapsulation forms: Payload type 14 is used for MPEG 1/2 audio, payload type 32 is used for MPEG 1/2 video elementary streams and payload type 33 is used for MPEG 1/2 transport streams [98]. When using RTP to carry MPEG encoded data, an MPEG specific header is inserted after the RTP header [63].

RTCP is used to provide feedback on the quality of the data transmission, and optionally to convey minimal session control information. It is based on periodic transmission of control packets. Sender reports and receiver reports are sent periodically to all the participants containing information relevant to the calculation of packet loss rate, packet transmission delay and delay jitter. These provide important information on the current state of the network. Some applications can use this information to adjust their transmission rates and to help relieve congestion.

RTP is used in all of the current MBone multimedia conferencing tools such as VAT, VIC and NV [75].

5.10 Evolution to IPv6

Internet Protocol Version 6 is the next generation Internet Protocol. The basic protocol description is given in Deering and Hinden's RFC1883 [43]; these authors present a technical overview in "Internet Protocol Next Generation" [62]. IPv6 has been designed largely to solve the address space crisis on the Internet. For details on IPv6 IP address allocation, see Rekhter and Li's RFC1887 [92]. It is a more streamlined protocol which cleans up many "relics" in IPv4 and also provides many new features such as address autoconfiguration (below), security [10], hierarchical routing (Nimrod), flow labeling, etc. The Flow Label Field [86] is not intended to be used in routing, though it is certainly considered as a forwarding label in Tag Switching. It together with the source and destination addresses identifies the flow to which a packet belongs. The label is necessary since the packet might be encrypted so that the port and protocol information is no longer available to routers.

Deployment of IPv6 in NBMA networks (and ATM networks in particular) presents a number of new challenges [11, 96]. Among the foremost is the use of address autoconfiguration (RFC1971, "IPv6 Stateless Address Autoconfiguration" [111]). Address autoconfiguration allows a host to configure one or more addresses per interface automatically and without explicit system administration. However, IPv6 fundamentally changed the Address Resolution Protocol. In IPv4, a different ARP protocol is defined for each medium. In IPv6, RFC1970 [77] defines a common Neighbor Discovery (ND) protocol for all media types. This ND protocol assumes that if the data link address of a certain node is not available, it still can be reached by sending a multicast message. This requires that the data link protocol inherently support multicast, which means there is a straightforward mapping between the IPv6 multicast address and the data link multicast address. Unfortunately, as we know, ATM's native multicast support is weak. It was originally proposed that the MARS model be used; routers would perform block MARS_JOINs for an appropriate range of IPv6 multicast addresses [5]. However, since the development of NHRP, ION has amended this strategy [11]; while MARS is used for ND, for destinations not considered as neighbors hosts send packets to their default router. The router in turn issues an NHRP query to determine the target's ATM address, and on learning it issues a Redirect to the IPv6 source, identifying the flow destination as a Transient Neighbor [9].

This is one of the stickier issues for bringing IPv6 to ATM networks. Routing models are much closer between IPv6 (Nimrod) and ATM (I–PNNI). However, IPv6 has not become as immediate an issue as we envisioned at the outset of this project. Deployment is in trial systems only, running over IPv4 tunnels. Alternative solutions to the "Address Explosion" problem which is IPv6's greatest *raison d'etre* have reduced the urgency with which IPv6's development was then regarded. Foremost among these is Classless Inter–Domain Routing (CIDR) [105]. Other limitations of IPv4 still drive the development and eventual deployment of IPv6, but not to any great degree within the horizons of this report.

6 Competing Technologies

Native ATM as a data transport medium, without IP, is certainly a "competing technology" to IP–ATM. A number of native protocol stacks have been produced, notably the native ATM support in Microsoft's Winsock and IBM's "Low Level ATM Interface" (LLATMI) [115]. However, we have decided to limit our overview to competing means of delivering IP (or other network–layer protocols), since this is the area of our experience. The primary alternative to IP over ATM is certainly "next generation" IP routers. The favored alternative for linking these routers together is SONET, with some form of point–to–point framing such as PPP. We will look at several alternative technologies in this section.

Peter Newman and others at Ipsilon Networks present an interesting if somewhat biased comparison of Gigabit router technology with IP Switching [82]. We discuss two of the same devices, MGBR and $a^I t^P m$, and two others, the Ascend GRF switch and the Pluris Massively Parallel Router.

6.1 Gigabit IP Routers - $a^I t^P m$

One of the few "Gigabit Routers" for which design details rather than simple schematic overviews have been published is the Washington University $a^I t^P m$ (that is, IP *over* ATM). Parulkar, Schmidt and Turner [87] describe a device which combines the attributes of the IP switch with those of Gigabit IP routers. It does this by using a custom non–blocking, multicast ATM switching fabric as the core of the router. Around the switched core are arrayed a number of routing cards and line interface cards. The routing cards are relatively simple, combining a CPU with PROM for program storage, RAM for CPU scratchpad, VRAM for packet buffers, and a custom–designed "ATM Port Interconnect Chip" (APIC). The APIC has Utopia (standard ATM-SONET) input and output interfaces for direct connection to host ATM interfaces, and a Sun Mbus for connection to the ATM switch fabric.

Each routing card can receive IP datagrams (on its incoming Utopia interface) which are PPP–encapsulated into AAL5 PDUs; this is the means by which these routers communicate with one another. It is also possible to construct line cards with Ethernet or other LAN interface, which perform the PPP and ATM AAL5 encapsulation on–board. The APIC moves the cells from this PDU into the VRAM buffer, the CPU (called the IP Processing Element, or IPPE) performs IP header processing, and directs a second APIC whether to send the datagram's cells to the outgoing Utopia interface or to the ATM switch core via the Mbus. The latter would be the case when the datagram was to be directed to another routing card, or to one of the line cards.

Firmware in the IPPE's PROM sets up a default Virtual Circuit to each other IPPE in the router (through the ATM core). This VC is used for communication and for forwarding datagrams. It requires header processing by both IPPEs. Header processing typically requires around 100 instructions. Each routing card also establishes a number of PVCs which link directly from APICS on incoming ports to APICs on outgoing ports. These are used, dynamically, for forwarding bursts of datagrams without further IPPE intervention; the criteria used for characterizing a "burst" (apparently similar to an IP Switch flow) is not discussed. PVCs from this pool are temporarily assigned to particular datagram streams, leading to a form of cut–through routing in which header processing after the initial datagram is not required. These PVCs are reclaimed when the burst ceases. Assigning a PVC to a burst is purely local within the router, and is estimated to require less than 100 microseconds per router traversed. The firmware is configured to handle RSVP messages, and one of the IPPEs functions as a Route Server IPPE for the entire router, running standard Internet routing protocols.

This design is a research tool, rather than a commercial router. It is interesting because, while it shares the basic approach of the IP Switch, it scales to a Gigabit router by implementing a number of routing cards in parallel, each with two OC-12c ports and the processing power to handle 1.2 Gbps throughput. Each line card has 12 OC-3c and one OC-12c ports. It appears that the net throughput of this router is limited only by the ATM switch core, not specified but certainly in the range of 10 Gbps or more. Unlike the IP Switch or Tag Switch, the $a^I t^P m$ does not need to communicate with adjacent routers to make a cut–through assignment.

6.2 Gigabit IP Routers - Ascend GRF IP Switch

While characterizing their product as an IP Switch, Ascend [39] produces a device very similar in operation to the $a^I t^P m$ router; internal details are far less well discussed, however, so it's difficult to be certain. Ascend has two versions of this switch, one based on a 4 Gbps non–blocking crossbar switch and the other on a 16 Gbps crossbar. The former allows up to 4 "media" cards, each with 1 Gbps capacity, while the latter accepts 16 media cards. Media cards are available with a variety of line interfaces: 100Base–T Ethernet, OC-3c ATM, OC-3c SONET with PPP and Frame Relay, FDDI, CDDI, or HSSI. Media cards perform next–hop lookup and forwarding, while a separate Route Manager (166 MHz Pentium) handles Internet routing protocols.

Each media card utilizes non–cached, hardware lookup in a 150,000 route entry table. It is claimed that the lookup time averages less than 1 microsecond, and is never more than 2.5 microseconds. It is not disclosed how datagrams are packaged for transport through the non–blocking switch core. The GRF switches gain their very great throughput from the distributed, parallel nature of the forwarding engines.

6.3 BBN's Multi–Gigabit Router

Known as the MGBR, BBN's Multi–Gigabit Router has been designed by Craig Partridge under a continuing grant from the Defense Advanced Research Projects Agency (DARPA) [100]. It may evolve into a commercial product but is far from that at present. Figure 14 illustrates the construction of this router. It is clearly very similar in spirit to the GRF switch. BBN is a manufacturer of crossbar switches; the one used here is of proprietary design but claimed to have the following properties:

- 64–bit wide data path (plus overhead bits)
- 52 MHz clock rate
- 3.3 Gbps full duplex throughput on each of 15 ports
- Aggregate 50 Gbps throughput to fabric

Each Forwarding Engine uses a 415 MHz Pentium and completes a 50 instruction cycle header lookup in about 120 ns; this translates to over 8 million packets per second forwarding rate. Line interfaces which are under development include HIPPI, FDDI, OC-12c and OC-48c ATM over SONET, though only the OC-12c ATM interface is intended to be supplied on the first routers.

6.4 Pluris Massively Parallel Router

This device is fairly similar to the above mentioned routers in that the work is done by a large number of simple "processing nodes" arranged around a switch core. In this instance, the core is a proprietary Self–Healing Butterfly Switch [3]. Each node utilizes IP framed on a single OC-3c SONET link; up to 16,000 processing nodes may be attached to the core switch, with their OC-3c links aggregated to higher rate SONET through SONET synchronous multiplexors. In this configuration (perhaps somewhat improbable) the router would be capable of 7 Gpps, or 2.4 Tbps!

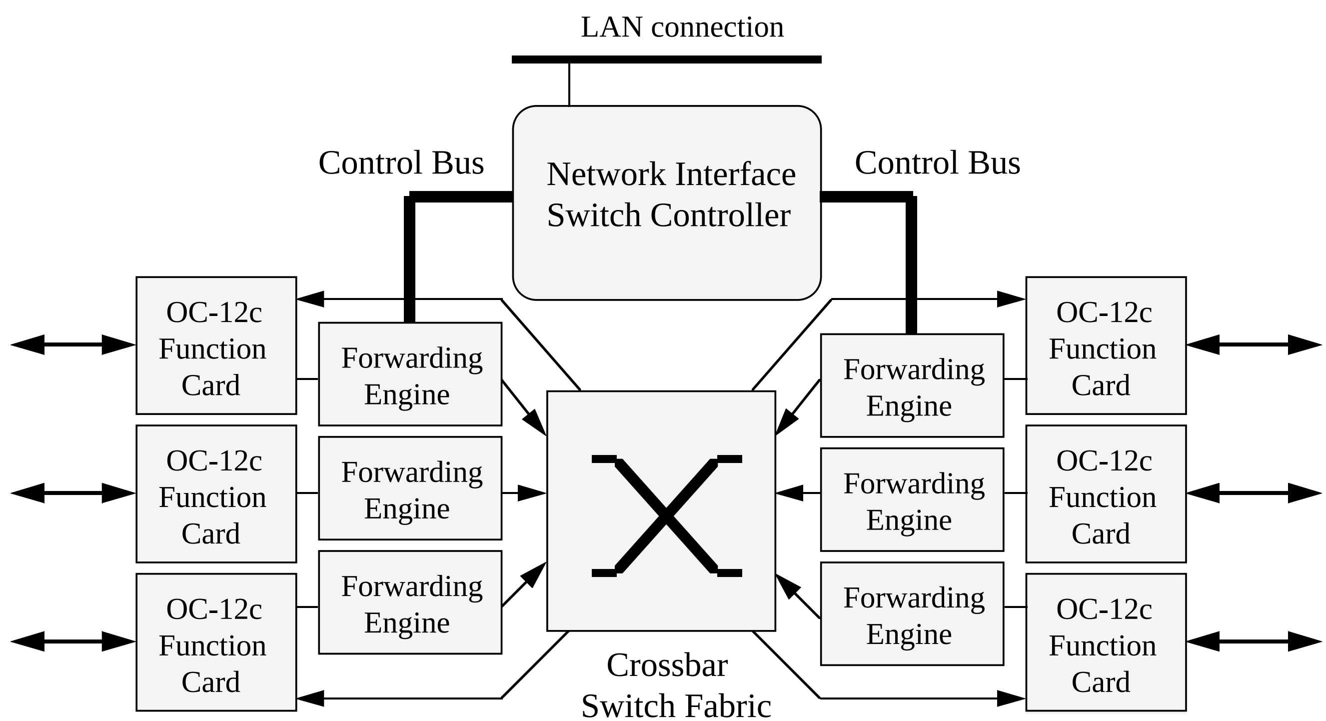


Figure 14: BBN-DARPA Multi-Gigabit Router

7 Performance Modeling of IP Over ATM

The end–to–end transport of an IP packet may involve the use of heterogeneous networking technologies. As an example, the packet may first be transmitted on an Ethernet, and then on an ATM backbone, and finally on another Ethernet before reaching its destination. The available networking technologies may have varying degrees of QoS support, e.g., CBR service from ATM and best–effort service from an Ethernet. Of interest is a detailed understanding of the interworking of potentially different QoS support, and the characterization of the resulting end–to–end QoS seen at the IP layer.

It is proposed to develop a simulation model for an IP network with possibly multiple underlying technologies. In particular, the following scenarios will be considered:

1. Underlying technology is ATM, with selected IP–over–ATM schemes.

2. Underlying technology is a mixture of ATM and Ethernet (shared or dedicated); same IP–over–ATM schemes as above.

For each IP–over–ATM scheme, our model takes into consideration features of routing and connection establishment/termination. Issues to be investigated include:

1. How to map QoS requirements at the IP layer to QoS provided by the underlying networking technologies.

2. The impact of dynamic variation of QoS parameters at the underlying networks on the QoS seen at the IP layer.

Attempts will be made to validate our simulation model using data obtained from performance monitoring. Recommendations on how best to utilize the various networking technologies to support end–to–end QoS at the IP layer, as well as the preferred IP–over–ATM scheme(s), will also be developed.

8 Conclusions

This was a difficult report to prepare, both because the technology and the options are changing rapidly, and because some of the issues have a decidedly "religious" overtone. What follows is our outlook on the primary issue of IP over ATM, and the derivative issue of video and audio streams over IP/ATM.

What is the future of ATM in data networking? From its beginnings as the multiplexing layer in B–ISDN, ATM's designers have presented it as the best hope for broadband, low delay, universally scalable data communications. However, it has been held in no such regard by Internet engineers, who find its design violates most of what they have learned about building robust and maintainable networks. As a result, even while some parts of both communities (telecom and datacom) have worked to make IP and other data communications protocols interwork successfully with ATM, others have rejected the idea vigorously, and are now developing alternative strategies for delivering real–time, high–speed data with useful quality guarantees.

So our most basic question is, *Will there be IP over ATM?* At some level, the answer is of course yes, certainly as long as backbone carriers find it economic to use ATM over SONET for connecting their routers. At present, virtually all do: MCI is using Nortel OC-12c and OC-48c ATM equipment for its data network; UUNet has upgraded its backbone to OC-12c ATM; CERFnet to OC-3c ATM; and the NSF's vBNS research network runs over MCI's OC-12c ATM. ATM is a circuit switching telco technology, and these carriers are basically telcos. However, backbone IP services over ATM are generally delivered using ATM for multiplexing and bandwidth sharing between routers, as a point–to–point link rather than through any of the more integrated techniques we have discussed. There have been many complaints within the Internet community about the bandwidth inefficiency of TCP/IP running over ATM. The TCP ACK, which is the most common packet (about 40% of IP traffic is TCP ACKs), uses only 40 data bytes; with its 23 bytes of LLC/SNAP encapsulation, it must be put into 2 ATM cells, leading to roughly 40% "efficiency" on ACKs and 78% for a "typical" traffic mix. While there *are* other alternatives to this grim scenario, it is one that is frequently quoted when it is claimed that backbone providers are using ATM only until something better is available at an equivalent price.

There are also strong business reasons for the continued use of ATM for delivery of data services in the WAN. Telcos have always had the goal of providing Value–Added services, rather than inexpensive pipelines. ATM offers a much finer granularity of bandwidth than is available from typical SONET muxes, and provisioned VCs between customer sites offer the opportunity for service differentiation. Backbone carriers have invested heavily in ATM switch technology, and need to recover their investments. For all these reasons, if telcos offer, for instance, raw SONET pipes to ISPs, the tariffs are likely to be prohibitive.

Nevertheless, the alternatives (in general, Gigabit IP routers or some form of Tag Switching) are drawing a great deal of interest, and a number of vendors will certainly offer such devices interfaced directly to SONET, Frame Relay, or other "connectionless" lower layers. Assuming that these devices turn out, when constructed (there are yet very few commercial Gigabit Routers, and no large–scale installations of which we are aware) to exhibit the speed, throughput, and scaling properties which are intended, the question is when, and whether, backbone providers will be willing to phase out their current investment in ATM equipment. There is also the open question of Quality of Service: Are QoS guarantees needed at all, and if so, can RSVP on IP routers deliver them?

In more local networks (MANs, Enterprise networks, Campus networks) we are likely to see a rapid introduction of non–ATM, high–speed IP router technology. Certainly within the traditional data communications communities, fast IP routers will win rapid acceptance. The investment in ATM equipment has been far lower, and the mindset is that of the Internet and of data communications, rather than switched services and telecommunications. If bandwidth becomes cheap enough, fast enough, the 3 tenets of "Best Effort" (Section 2) will be realized and Gigabit (Terabit, etc.) routers will push ATM out of the Intranet.

The third area of major investment in new Internet carrier equipment is in cable data networks, particular Hybrid Fiber–Cable (HFC). Many or most of these (for instance, the @Home service provider network) use a private ATM network with NAPs for access to the Internet, and ATM switches in the cable head–ends for data service to customers. However, we have found it very difficult to learn how these networks use ATM; whether just as a convenient physical medium to carry IP between routers and customers, or an integral part of their data communications network. Hence we find it impossible to say how the availability of a new generation of IP routers will affect their architectural plans.

Insofar as ATM remains a presence in the Internet, in Intranets, and in local data delivery, there is a certainty of many interesting protocol interactions between multimedia applications, network, and datalink layers. Encapsulation of MPEG TS or PS into IP datagrams, then into ATM cells, will require careful analysis if any sort of efficiency is going to be attained. If the dominant model becomes Best Effort, then adaptive applications will be mandatory. We expect that multicast will grow rapidly in importance, both for communication and entertainment in the Internet, and for applications such as software update and data distribution in Intranets. However, while multicast is an almost indispensable tool in LANs, the tools to use it effectively in a wider area are only being developed, and there is an element of uncertainty in this expectation. Assuming ATM continues to grow in its installed base, then a multicast requirement will likely drive the choice of IP–ATM models toward MPOA. Again, though, the first practical tests of the technology are just beginning. IPv6, I–PNNI, these are technologies further in the future than was apparent 6 months or a year ago, and too far for us to attempt any estimate of success or outcome.

We live in interesting times!

A Active IETF Working Groups

This list and the associated charters were generated on Wednesday 28 May 1997 at 10:51:04.

A.1 IETF Areas

- Applications Area
- General Area
- Internet Area
- Operations and Management Area
- Routing Area
- Security Area
- Transport Area
- User Services Area

A.2 Areas and Their Working Groups

Applications Area

Area Director(s)

- Keith Moore `<moore@cs.utk.edu>`
- Harald Alvestrand `<Harald.T.Alvestrand@uninett.no>`

Working groups:

- Access, Searching and Indexing of Directories (asid)
- Application Configuration Access Protocol (acap)
- Application MIB (applmib)
- Calendaring and Scheduling (calsch)
- Common Indexing Protocol (find)
- Detailed Revision/Update of Message Standards (drums)

- Electronic Data Interchange-Internet Integration (ediint)
- Extensions to FTP (ftpext)
- HyperText Transfer Protocol (http)
- Integrated Directory Services (ids)
- Internet Fax (fax)
- Internet Printing Protocol (ipp)
- Large Scale Multicast Applications (lsma)
- MIME - X.400 Gateway (mixer)
- MIME Encapsulation of Aggregate HTML Documents (mhtml)
- Mail and Directory Management (madman)
- NNTP Extensions (nntpext)
- Printer MIB (printmib)
- Receipt Notifications for Internet Mail (receipt)
- Telnet TN3270 Enhancements (tn3270e)
- Uniform Resource Names (urn)
- WWW Distributed Authoring and Versioning (webdav)

General Area

Area Director(s)

- Fred Baker <fred@cisco.com>

Working groups:

- Process for Organization of Internet StandardS ONg (poisson)

Internet Area

Area Director(s)

- Thomas Narten <narten@raleigh.ibm.com>
- Jeffrey Burgan <burgan@home.net>

Working groups:

- 100VG-AnyLAN MIB (vgmib)
- AToM MIB (atommib)
- DNS IXFR, Notification, and Dynamic Update (dnsind)
- DS1/DS3 MIB (trunkmib)
- Dynamic Host Configuration (dhc)
- Frame Relay Service MIB (frnetmib)
- IP over Cable Data Network (ipcdn)
- IPNG (ipngwg)
- IPv6 MIB (ipv6mib)
- ISDN MIB (isdnmib)
- Interfaces MIB (ifmib)
- Internetworking Over NBMA (ion)
- PacketWay (pktway)
- Point-to-Point Protocol Extensions (pppext)
- Service Location Protocol (svrloc)

Operations and Management Area

Area Director(s)

- John Curran `<jcurran@bbn.com>`
- Michael O'Dell `<mo@uu.net>`

Working groups:

- Benchmarking Methodology (bmwg)
- Bridge MIB (bridge)
- Distributed Management (disman)
- Entity MIB (entmib)
- G and R for Security Incident Processing (grip)
- Guide for Internet Standards Writers (stdguide)
- IEEE 802.3 Hub MIB (hubmib)
- MBONE Deployment (mboned)
- New Generation Transition (ngtrans)
- Physical Topology MIB (ptopomib)
- Procedures for Internet/Enterprise Renumbering (pier)
- RWhois Operational Development (rwhois)
- Realtime Traffic Flow Measurement (rtfm)
- Remote Authentication Dial-In User Service (radius)
- Remote Network Monitoring (rmonmib)
- Roaming Operations (roamops)
- Routing Policy System (rps)
- SNMP Agent Extensibility (agentx)
- SNMP Version 3 (snmpv3)
- The Internet and the Millennium Problem (2000)
- Uninterruptible Power Supply (upsmib)

Routing Area

Area Director(s)

- Joel Halpern `<jhalpern@newbridge.com>`

Working groups:

- Data Link Switching MIB (dlswmib)
- IP Routing for Wireless/Mobile Hosts (mobileip)
- IS-IS for IP Internets (isis)
- Inter-Domain Multicast Routing (idmr)
- Inter-Domain Routing (idr)
- Multicast Extensions to OSPF (mospf)
- Multiprotocol Label Switching (mpls)
- New Internet Routing and Addressing Architecture (nimrod)
- Open Shortest Path First IGP (ospf)
- QoS Routing (qosr)
- Routing Information Protocol (rip)
- SNA DLC Services MIB (snadlc)
- SNA NAU Services MIB (snanau)
- Source Demand Routing (sdr)
- UniDirectional Link Routing (udlr)

Security Area

Area Director(s)

- Jeffrey Schiller `<jis@mit.edu>`

Working groups:

- Authenticated Firewall Traversal (aft)
- Common Authentication Technology (cat)
- Domain Name System Security (dnssec)
- IP Security Protocol (ipsec)

- One Time Password Authentication (otp)
- Public-Key Infrastructure (X.509) (pkix)
- Secure Shell (secsh)
- Simple Public Key Infrastructure (spki)
- Transport Layer Security (tls)
- Web Transaction Security (wts)

Transport Area

Area Director(s)

- Scott Bradner `<sob@harvard.edu>`
- Allyn Romanow `<allyn.romanow@eng.sun.com>`

Working groups:

- Audio/Video Transport (avt)
- Integrated Services (intserv)
- Integrated Services over Specific Link Layers (issll)
- Multiparty Multimedia Session Control (mmusic)
- ONC Remote Procedure Call (oncrpc)
- Resource Reservation Setup Protocol (rsvp)
- TCP Implementation (tcpimpl)
- TCP Large Windows (tcplw)

User Services Area

Area Director(s)

- Joyce Reynolds `<jkrey@isi.edu>`

Working groups:

- Humanities and Arts (harts)
- Internet School Networking (isn)
- Responsible Use of the Network (run)
- Site Security Handbook (ssh)
- User Services (uswg)

B ITU Study Groups

Study Group 2 Network and service operation

Study Group 3 Tariff and accounting principles

Study Group 4 TMN and network maintenance

Study Group 5 Protection against electromagnetic environment effects

Study Group 6 Outside plant

Study Group 7 Data networks and open system communications

Study Group 8 Characteristics of telematic systems

Study Group 9 Television and sound transmission

Study Group 10 Languages and general software aspects for telecommunication systems

Study Group 11 Signalling requirements and protocols

Study Group 12 End-to-end transmission performance of networks and terminals

Study Group 13 General network aspects including GII

Study Group 15 Transport networks, systems and equipment

Study Group 16 Multimedia services and systems

Source: ITU Web Site: `<http://www.itu.ch>`

C ATM Forum Working Groups, Specification and Contributions

C.1 ATM Forum contributions

The following Contributions reflect the current work of the Technical Working Groups, which will lead to new or revised Specifications.

94-0471 PNNI Draft Specification (R13)
Editor: Doug Dykeman
Chair: Mike Goguen

96-0314 RSVP Birds of a Feather Meeting Minutes, February 1996
Author: Caralyn Brown

96-0353 Straw Ballot Comments for 155.52 Mbps Short Wavelength Laser Physical Layer Specification (ATM Forum/95-1301R3)
Author: Robert R. Campbell

96-0354 An Overview of PNNI Augmented Routing
Author: Ross Callon
Jason Jeffords
John Drake
Hal Sandick
Joel Halpern

96-0355 Issues and Approaches for Integrated PNNI
Author: Ross Callon
Jason Jeffords
Hal Sandick
Joel Halpern

C.2 Working Groups and Specifications

The following Table 4, Table 5, Table 6 and Table 7 list the ATM Forum's Technical Working Groups and the specifications they have developed and approved as of December 1996; it is the most current version available as of May 1997, however.

Source: `<http://www.atmforum.com/>`

Last updated: Mon Apr 8 11:45:49 PDT 1997

Table 4: ATM Forum Spec Watch: Approved Items as of December 1996

Technical Working Group	Approved Specifications	Specification
B-ICI	B-ICI 1.0 B-ICI 1.1 B-ICI 2.0 (delta spec to B-ICI 1.1) B-ICI 2.0 (integrated specification) B-ICI 2.0 (Addendum or 2.1)	af-bici-0013.000 af-bici-0013.001 af-bici-0013.002 af-bici-0013.003 af-bici-0068.000
Data Exchange Interface	Data Exchange Interface version 1.0	af-dxi-0014.000
Integrated Layer Management Interface	ILMI 4.0	af-ilmi-0065.000
LAN Emulation	LAN Emulation over ATM 1.0 LAN Emulation Client Management Specification LANE 1.0 Addendum LANE Servers Management Spec v1.0	af-lane-0021.000 af-lane-0038.000 af-lane-0050.000 af-lane-0057.000
Network Management	Customer Network Management (CNM) for ATM Public Network Service M4 Interface Requirements and Logical MIB CMIP Specification for the M4 Interface M4 Public Network view	af-nm-0019.000 af-nm-0020.000 af-nm-0027.000 af-nm-0058.000
P-NNI	Interim Inter-Switch Signaling Protocol P-NNI V1.0 P-NNI V1.0 Addendum	af-pnni-0026.000 af-pnni-0055.000 af-pnni-0066.000
Signaling	(See UNI 3.1, af-uni-0010.002) UNI Signaling 4.0	af-sig-0061.000
Traffic Management	(See UNI 3.1, af-uni-0010.002)	

Table 5: ATM Forum Spec Watch: Approved Items as of December 1996

Technical Working Group	Approved Specifications	Specification
Physical Layer	Issued as part of UNI 3.1: 44.736 DS3 Mbps Physical Layer 100 Mbps Multimode Fiber Interface Physical Layer 155.52 Mbps SONET STS-3c Physical Layer 155.52 Mbps Physical Layer	af-uni-0010.002
	ATM Physical Medium Dependent Interface Specification for 155 Mb/s over Twisted Pair Cable	af-phy-0015.000
	DS1 Physical Layer Specification Utopia Level 1 v2.01 Mid-range Physical Layer Specification for Category 3 UTP	af-phy-0016.000 af-phy-0017.000 af-phy-0018.000
	6,312 Kbps UNI Specification E3 UNI Utopia Level 2 v1.0 Physical Interface Specification for 25.6 Mb/s over Twisted Pair	af-phy-0029.000 af-phy-0034.000 af-phy-0039.000 af-phy-0040.000
	A Cell-based Transmission Convergence Sublayer for Clear Channel Interfaces	af-phy-0043.000
	622.08 Mbps Physical Layer 155.52 Mbps Physical Layer Specification for Category 3 UTP (See also UNI 3.1, af-uni-0010.002)	af-phy-0046.000 af-phy-0047.000
	120 Ohm Addendum to ATM PMD Interface Spec for 155 Mbps over TP	af-phy-0053.000
	DS3 Physical Layer Interface Spec 155 Mbps over MMF Short Wave Length Lasers, Addendum to UNI 3.1	af-phy-0054.000 af-phy-0062.000
	WIRE (PMD to TC layers) E-1	af-phy-0063.000 af-phy-0064.000

Table 6: ATM Forum Spec Watch: Approved Items as of December 1996

Technical Working Group	Approved Specifications	Specification
Service Aspects and Applications	Frame UNI	af-saa-0031.000
	Circuit Emulation	af-saa-0032.000
	Native ATM Services: Semantic Description	af-saa-0048.000
	Audio/Visual Multimedia Services: Video on Demand Specification	af-saa-0049.000
	ATM Names Service	af-saa-0069.000
User-Network Interface (UNI)	ATM User-Network Interface Specification V2.0	af-uni-0010.000
	ATM User-Network Interface Specification V3.0	af-uni-0010.001
	ATM User-Network Interface Specification V3.1	af-uni-0010.002
	ILMI MIB for UNI 3.0	af-uni-0011.000
	ILMI MIB for UNI 3.1	af-uni-0011.001
	Traffic Management 4.0	af-tm-0056.000

Table 7: ATM Forum Spec Watch: Approved Items as of December 1996

Technical Working Group	Approved Specifications	Specification
Testing	Introduction to ATM Forum Test Specifications	af-test-0022.000
	PICS Proforma for the DS3 Physical Layer Interface	af-test-0023.000
	PICS Proforma for the SONET STS-3c Physical Layer Interface	af-test-0024.000
	PICS Proforma for the 100 Mbps Multimode Fibre Physical Layer Interface	af-test-0025.000
	PICS Proforma for the ATM Layer (UNI 3.0)	af-test-0028.000
	Conformance Abstract Test Suite for the ATM Layer for Intermediate Systems (UNI 3.0)	af-test-0030.000
	Interoperability Test Suite for the ATM Layer (UNI 3.0)	af-test-0035.000
	Interoperability Test Suites for Physical Layer: DS-3, STS-3c, 100 Mbps MMF (TAXI)	af-test-0036.000
	PICS for DS-1 Physical Layer	af-test-0037.000
	Conformance Abstract Test Suite for the ATM Layer (End Systems) UNI 3.0	af-test-0041.000
	PICS for AAL5 (ITU spec)	af-test-0042.000
	PICS Proforma for the 51.84 Mbps Mid-Range PHY Layer Interface	af-test-0044.000
	Conformance Abstract Test Suite for the ATM Layer of Intermediate Systems (UNI 3.1)	af-test-0045.000
	PICS for the 25.6 Mbps over Twisted Pair Cable (UTP-3) Physical Layer	af-test-0051.000
	PICS for ATM Layer (UNI 3.1)	af-test-0059.000
	Conformance Abstract Test Suite for the UNI 3.1 ATM Layer of End Systems	af-test-0060.000
	Conformance Abstract Test Suite of the SSCOP for UNI 3.1	af-test-0067.000
	PICS for the 155 Mbps over Twisted Pair Cable (UTP-5/STP-5) Physical Layer	af-test-0070.000

D Alternative Media and Encapsulations

D.1 Cells In Frames (CIF)

"Cells In Frames" is a concept introduced by Grenville Armitage at least as long ago as 1990 [8] but which has recently enjoyed a resurgence of interest in the industry. This is due to the work of a collaboration between Richard Cogger and Scott Brim of Cornell University, and Larry Roberts of Connectware, but has evinced interest by such vendors as 3Com, IBM, Apple, Microsoft, PMC-Sierra, Stratacom, and Sun Microsystems [117].

CIF is a means of getting a native ATM API to the desktop using existing physical and datalink layer devices, particularly Ethernet v2 and 802.3, and 802.5 Token Ring NICs. It is intended that this will provide an impetus for early experience with the benefits of native ATM as opposed to IP over ATM, particularly its promise of true QoS guarantees. It also provides an avenue for organizations which wish to take advantage of an ATM backbone using technologies such as the Ipsilon IP Switch, to integrate existing workstations using legacy LAN hardware directly into the system.

The basis of CIF is the specification of a header and framing format for carrying ATM cells within Ethernet/802.3 or Token Ring 802.5 frames. This specification is now in Version 1.0 [22]. Two frame formats (Format 0 and 1) are defined for communication between end-station and attachment device; such communication is out-of-band with respect to ATM and is used in setting up and managing local communication only. The major format type (Format 2) encapsulates ATM cells. ATM cells are arranged in one or more groups, each group representing a single Virtual Circuit (VC). Cells within a group carry only a single 5-byte ATM header for all following 48-octet SAR-PDUs, and one Ethernet/802.3 frame can carry up to 31 cells so packaged. At the other extreme, each cell an independent group with its own header, 26 cells will fit into the Ethernet/802.3 frame. While the 802.5 Token Ring specification supports a much larger frame size (4500 bytes, but up to 17Kbytes for 16Mbps Token Ring), at present the CIF specification does not take advantage of this.

The CIF specification also includes means of relieving the host of some of the load of ATM and Adaptation Layer processing. This is based on the idea that networks will introduce CIF Attachment Devices (CIF-AD) which are more than legacy LAN hubs or even LAN switching devices. A CIF–AD would typically be an ATM workgroup switch with Ethernet or Token Ring ports, having full ATM signalling, ILMI, and OAM support. The CIF setup protocol provides a means for the end-station to discover whether the CIF–AD is capable of performing ABR traffic management, AAL5 CRC–32 calculation, and cell segmentation–and–reassembly, thus relieving the end–station of these CPU–intensive tasks. It is likely that much of the industry interest in this proposal stems from the opportunity to develop VLSI and attachment devices in support of its adoption. In light of the growing volume of 100Mbps Ethernet installations, there may well be a niche for this technology.

D.2 Residential Broadband Services

Cable modems are one aspect of the very rapid and well-financed provision of residential broadband services. The goals are to provide access to entertainment, interactive television, telephone, network and internetwork services. Both telcos and cable networks are converging on this arena. There are 4 contending Residential Access Network (RAN) architectures [65]:

- HFC (Hybrid Fiber–Coax)
- FTTC (Fiber–to–the–Curb)
- FTTH (Fiber–to–the–Home)
- ADSL (Asynchronous Digital Subscriber Loop)

Cable companies are spending large sums (estimated to be many billions of dollars) to upgrade physical plant from passive coaxial cable trees to Hybrid Fiber-Coax (HFC) networks, just to keep pace with the demand for ordinary cable TV service delivery. Time-Warner alone, for instance, is spending at least $4 billion [57] in its current round of infrastructure expansion. In HFC cable systems, signal distribution is via fiber optics from the head-end to neighborhood nodes, where bridging devices convert the signals for delivery to the end-user on standard coaxial cable. HFC is still a broadcast, passband technology, ideal for conventional television signal delivery. In order to support digital services, digital modulation is required; QAM and VSB are the most common. Cable modems supporting these techniques are capable of carrying as much as 36Mbps of user data over a 6MHz channel. Still unresolved is the issue of which link layer will predominate. There is significant support in the cable industry for both MPEG-2 Transport System (MPEG-TS) and for native ATM in this role. MPEG-TS has the advantage of familiarity, inexpensive silicon for QAM-encoding, and slightly higher efficiency when contiguous ATM cells are packaged into 188 TS frames than when TS frames are carried in ATM cells. The efficiency arguement for this "Cells In Frames"-like arrangement is rather misleading, however, in that the supposed inefficiency of ATM cells is due to the presence of cell headers; these are not just a waste of bandwidth, but rather contribute greatly to ATM's flexibility in switching and signalling, QoS guarantees, and the possibility of an end-to-end Residential Broadband Network without low-level gateways.

For these reasons and others, the Advanced Television Systems Committee, an industry standards group, favours the MPEG-TS over ATM approach [1]. However, DAVIC has adopted MPEG-TS as the link layer of choice. This will no doubt lead to a mix of systems in cable networks. Telcos, on the other hand, have almost universally chosen ATM as their broadband platform for all delivery systems, for data services, multimedia, and their entertainment offerings [42].

Fiber–to–the–Curb, also known as Switched Digital Video, makes use of fiber optic connections terminating in Optical Network Units (ONU) "at the curb", servicing small groups of homes. Connections from the ONU to individual homes are made with twisted pair and coax. This is a baseband technology and targeted at telephony and digital interactive services. While TDM is used at present to multiplex signals for delivery to individual homes, ATM switches will likely replace the ONU in the near future [65].

Fiber–to–the–Home trials are being implemented by several telcos, notably NTT and Deutsche Telekom. Intrinsically a point–to–point network, this is a natural vehicle for ATM delivery; bandwidth can run to OC-3 levels. In initial installations, however, passive splitters are used to serve several homes economically, necessitating a MAC protocol for sharing bandwidth and access; this requires the development of "enhancements" to ATM for upstream medium sharing.

ADSL has the lowest performance of the contending methods, but uses standard twisted pair to support as much as 1.5Mbps downstream and 64kbps upstream. This makes it a straightforward entry vehicle for telcos. This bandwidth is insufficient for MPEG-2; this has led to the definition of ADSL-2, with a downstream bandwidth of 6Mbps and an upstream bandwidth of 640kbps. As this is a more expensive technology, it has not been widely deployed.

While cable companies have concentrated on HFC installation, the remaining 3 technologies are primarily telco-oriented. ADSL is still the easiest for conventional telcos to install, and with the adoption of new ATM physical layer interfaces can be used for the delivery of both digital services and digital video to the home. However, low intrinsic bandwidth limits its potential and expansion capability.

Meanwhile, manufacturers of cable system head-end equipment such as Com21 have begun to use ATM on the fiber distribution system for all service delivery, and are replacing analog head-ends with ATM-based head-end channel switches. This favors the use of ATM for data services as well as for video transport, leading to quicker development of ATM cable modems. Com21, a 3Com partner, supplies ComPORT ATM cable modems at present with Ethernet (10BaseT) connections to end-stations. However, companies such as Texas Instruments are working on minimal chipsets or even single-chip implementations of ATM cable modems, for direct installation on computer motherboards and in set-top boxes. As these become available, it seems likely that ATM will become the standard delivery technology in the cable TV industry, both for data and for television and multimedia entertainment. Fore Systems and General Instruments have been collaborating for some time on the development of end-to-end ATM networking for HFC. Fore Systems is contributing the head-end switching products and 25Mbps ATM host adapters, GI is responsible for the cable modem and applications, while the head-end processor is a joint development project [18].

Time-Warner is working with Com21 and 3Com to promote open standards for transmitting data over cable. The standardization efforts to date have not led to a consistent and interoperable set of either physical layer devices or protocols. However, given that the size of the "cable" market is far greater than that of ATM in traditional network and internetwork settings, it demands serious attention.

E List of Acronyms

AAL – **A**TM **A**daptation **L**ayer

AAL5 – **AAL** version **5** *(optimized for data transport)*

ABR – **A**vailable **B**it **R**ate

ADSL – **A**symmetric **D**igital **S**ubscriber **L**ine

API – **A**pplication **P**rogramming **I**nterface

ARIS – **A**ggregate **R**oute-based **I**P **S**witching

ARP – **A**ddress **R**esolution **P**rotocol

ATM – **A**synchronous **T**ransfer **M**ode

ATMARP – **ATM** **A**ddress **R**esolution **P**rotocol *(Classical IP over ATM service)*

ATSC – **A**dvanced **T**elevision **S**ystems **C**ommittee

ATTH – **A**TM **T**o **T**he **H**ome

BGP – **B**order **G**ateway **P**rotocol

B-ICI – **B**–ISDN **I**nter–**C**arrier **I**nterface

B-ISDN – **B**roadband – **ISDN**

BUS – **B**roadcast and **U**nknown **S**erver *(LANE)*

CAC – **C**onnection **A**dmission **C**ontrol

CBR – **C**onstant **B**it **R**ate

CDDI – **C**opper–**D**istributed **D**ata **I**nterface *(FDDI over UTP)*

CIDR – **C**lassless **I**nter–**D**omain **R**outing

CIF – **C**ells **I**n **F**rames

CIF–AD – **CIF** **A**ttachment **D**evice

CMFS – **C**ontinuous **M**edium **F**ile **S**erver

COS – **C**ommercial **O**ff–the–**S**helf *(Military purchasing jargon)*

CRC-32 – **C**yclic **R**edundancy **C**heck – **32** bits

CSR – **C**ell–**S**witching **R**outer

DAVIC – **D**igital **A**udio–**V**isual **C**ouncil

DNS – **D**omain **N**ame **S**erver

DSM–CC – **D**igital **S**erver **M**edia—**C**ommand and **C**ontrol

DTL – **D**esignated **T**ransit **L**ist *(P–NNI)*

DVMRP – **D**istance **V**ector **M**ulticast **R**outing **P**rotocol

ELAN – **E**mulated **L**ocal **A**rea **N**etwork

EPFL – **É**cole **P**olytechnique **F**édérale de **L**ausanne

FDDI – **F**iber–**D**istributed **D**ata **I**nterface

FIB – **F**orwarding **I**nformation **B**ase *(Cisco Tag Switching)*

FTP – **F**ile **T**ransfer **P**rotocol

FTTC – **F**iber **T**o **T**he **C**urb

FTTH – **F**iber **T**o **T**he **H**ome

GCAC – **G**eneric **C**onnection **A**dmission **C**ontrol

GSMP – **G**eneric **S**witch **M**anagement **P**rotocol *(Ipsilon IP Switch)*

HFC – **H**ybrid **F**iber–**C**oaxial Cable Network

HTTP – **H**yper **T**ext **T**ransfer **P**rotocol

IAB – **I**nternet **A**rchitecture **B**oard

IDMR – **I**nter **D**omain **M**ulticast **R**outing

IEEE – **I**nstitute of **E**lectrical and **E**lectronics **E**ngineers, Inc.

IETF – **I**nternet **E**ngineering **T**ask **F**orce

IFMP – **I**psilon **F**low **M**anagement **P**rotocol *(Ipsilon IP Switch)*

ILMI – **I**ntegrated **L**ocal **M**anagement **I**nterface *(LANE)*

INTSERV – **INT**egrated **SERV**ices *(IETF Working Group)*

ION – **I**P **O**ver **N**on–Broadcast Multiple Access Networks *(IETF Working Group)*

IOS – **I**nternetwork **O**perating **S**ystem *(Cisco Systems)*

IP – **I**nternet **P**rotocol

IPATM – **IP** Over **ATM** *(IETF Working Group)*

IPNG – **IP** **N**ext–**G**eneration *(IETF Working Group)*

I-PNNI – **I**ntegrated **PNNI**

IPv4 – **I**nternet **P**rotocol **v**ersion **4**

IPv5 – **I**nternet **P**rotocol **v**ersion **5** *(ST2+)*

IPv6 – **I**nternet **P**rotocol **v**ersion **6** *(IPNG)*

ISDN – **I**ntegrated **S**ervices **D**igital **N**etwork

ISSLL – **I**ntegrated **S**ervices Over **S**elected **L**ower **L**ayers *(IETF Working Group)*

InATMARP – **In**verse **ATMARP**

ITU – **I**nternational **T**elecommunications **U**nion

JPEG – **J**oint **P**hotographic **E**xperts **G**roup

LAG – **L**ocal **A**ddress **G**roup *(NHRP)*

LAN – **L**ocal **A**rea **N**etwork

LANE – **L**ocal **A**rea **N**etwork **E**mulation *(ATM Forum)*

LE_ARP – **L**AN **E**mulation **ARP** Server *(LANE)*

LEC – **L**AN **E**mulation **C**lient *(LANE)*

LES – **L**AN **E**mulation **S**erver *(LANE)*

LIJ – **L**eaf–**I**nitiated **J**oin

LIS – **L**ocal **I**P **S**ubnet

LLATMI – **L**ower–**L**ayer **ATM** **I**nterface

LLC – **L**ogical **L**ink **C**ontrol

LLC/SNAP – **LLC / S**ub **N**etwork **A**ccess **P**rotocol

LNNI – **L**AN-Emulation **N**etwork–**N**ode **I**nterface

MAC – **M**edium **A**ccess **C**ontrol

MARS – **M**ulticast **A**ddress **R**e**S**olution Protocol

MARSMCS – **M**ulticast **A**ddress **R**e**S**olution Protocol **M**ulti**C**ast **S**erver

MBone – **M**ulticast Back**Bone**

MLIS – **M**ulticast **L**ocal **I**P **S**ubnet

MOSPF – **M**ulticast **O**pen **S**hortest **P**ath **F**irst Routing Protocol

MPEG – **M**otion **P**icture **E**xperts **G**roup *(video encoding specification)*

MPEG–TS – **MPEG**—**T**ransport **S**tream

MPOA – **M**ulti–**P**rotocol **O**ver **A**TM

NAP – **N**etwork **A**ccess **P**oint

NBMA – **N**on–**B**roadcast **M**ultiple **A**ccess Media

ND – **N**eighbour **D**iscovery

NHC – **N**ext **H**op Resolution Protocol **C**lient *(NHRP)*

NHRP – **N**ext **H**op **R**esolution **P**rotocol

NHS – **N**ext **H**op Resolution Protocol **S**erver *(NHRP)*

NIC – **N**etwork **I**nterface **C**ard

NNI – **N**etwork–**N**ode **I**nterface

NSAP – **N**etwork **S**ervice **A**ccess **P**oint

NTP – **N**etwork **T**ime **P**rotocol

NV – **N**etwork **V**ideo Tool *(Xerox PARC)*

OAM – **O**perations, **A**dministration and **M**aintenance *(ATM Adaptation Layer administrative functions*

OC–3 – **O**ptical **C**arrier–**3** *(3 X 51.84 Mbps line rate)*

OSPF – **O**pen **S**hortest **P**ath **F**irst Routing Protocol

ONU – **O**ptical **N**etwork **U**nit

P2MP – **P**oint **2** (To) **M**ulti**P**oint

PATH – **PATH** Establishment Message *(RSVP)*

PDU – **P**rotocol **D**ata **U**nit *(packet)*

PGL – **P**eer **G**roup **L**eader *(P–NNI)*

PHY – **PHY**sical Layer Interface

PIM – **P**rotocol–**I**ndependent **M**ulticast

PIM-DM – **P**rotocol–**I**ndependent **M**ulticast–**D**ense **M**ode

PIM-SM – **P**rotocol–**I**ndependent **M**ulticast–**S**parse **M**ode

P–NNI – **P**rivate–**N**etwork–**N**ode **I**nterface *or alternatively, Private Network to Network Interface*

PPP – **P**oint To **P**oint **P**rotocol

PSTN – **P**ublic **S**witched **T**elephone **N**etwork

PTSE – **P**–NNI **T**opology **S**tate **E**lement

PVC – **P**ermanent **V**irtual **C**ircuit *(ATM)*

PVP – **P**ermanent **V**irtual **P**ath *(ATM)*

Q.2931 – *ITU–TS Signalling Specification for ATM in public networks*

QAM – **Q**uadrature **A**mplitude **M**odulation

QoS – **Q**uality **o**f **S**ervice

RBB – **R**esidential **B**road**B**and

RESV Resource **RES**er**V**ation Message *(RSVP)*

RFC – **R**equest **F**or **C**omment *(IETF specification)*

ROLC – **R**outing **O**ver **L**arge **C**louds *(IETF ATM Routing Working Group)*

RSVP – **R**esource Re**S**er**V**ation **P**rotocol

RTCP – **R**eal–**T**ime **C**ontrol **P**rotocol *(RTP)*

RTP – **R**eal–**T**ime **P**rotocol *(multimedia transport)*

SAR – **S**egmentation **A**nd **R**eassembly

SAR-PDU – **SAR–P**rotocol **D**ata **U**nit *(ATM cell data envelope)*

SDH – **S**ynchronous **D**igital **H**ierarchy

SMART – **S**hared **M**any–to–Many **A**TM **R**eserva**T**ions

SMDS – **S**witched **M**ultimegabit Per Second **D**ata **S**ervices

SMTP – **S**imple **M**ail **T**ransfer **P**rotocol *(email)*

SNMP – **S**imple **N**etwork **M**anagement **P**rotocol

SONET – **S**ynchronous **O**ptical **NET**work

SONET/SDH – **SONET/S**ynchronous **D**igital **H**ierarchy

SPANS – **S**imple **P**rotocol for **A**TM **N**etwork **S**ignalling *(Fore Systems)*

SSCOP – **S**ervice **S**pecific **C**onnection **O**riented **P**rotocol

ST2 – Internet **ST**ream Protocol Version **2**

SVC – **S**witched **V**irtual **C**ircuit

SVP – **S**witched **V**irtual **P**ath

T1 – **T**ime–Division Multiplexed Carrier **1** *(1 X 1.544 Mbps line rate)*

TDM – **T**ime **D**ivision **M**ultiplex

TCP – **T**ransmission **C**ontrol **P**rotocol

TDP – **T**ag **D**istribution **P**rotocol *(Cisco tag switching)*

TIB – **T**ag **I**nformation **B**ase *(Cisco tag switching)*

UDP – **U**ser **D**atagram **P**rotocol

UNI – **U**ser–**N**etwork **I**nterface

UNI3.0/3.1/4.0 – **UNI** version **3.0/3.1/4.0** *(ATM Forum)*

VAT – **V**isual **A**udio **T**ool *(LBL)*

VC – **V**irtual **C**hannel

VC – **V**irtual **C**ircuit

VCI – **V**irtual **C**hannel **I**dentifier

VCI – **V**irtual **C**ircuit **I**dentifier

VCI – **V**irtual **C**onnection **I**dentifier

VIC – **VI**deo **C**onference Tool *(LBL)*

VLAN – **V**irtual **L**ocal **A**rea **N**etwork

VLSI – **V**ery **L**arge **S**cale **I**ntegrated Circuit

VP – **V**irtual **P**ath

VPI – **V**irtual **P**ath **I**dentifier

VSB – **V**estigial **S**ide**B**and Modulation

WB – Shared **W**hite**B**oard Tool *(LBL)*

F Bibliography

Bibliography

[1] Advanced Television Systems Committee. *Guide to the Use of the ATSC Digital Television Standard*, doc. A/54 edition, Oct 1995.

[2] Anthony Alles. ATM internetworking. Cisco Document #030802 4/95 GP, <http://www.cisco.com/warp/public/614/atm_wp.pdf>, May 1995.

[3] Vadim Antonov. Pluris Massively Parallel Routing. Pluris Web Page, <http://www.pluris.com/wp/index.html>, Feb 1997.

[4] Grenville J. Armitage. VENUS - Very Extensive Non-Unicast Service. IETF draft-armitage-ion-venus-02.txt, 04/22/1997.

[5] Grenville J. Armitage. IPv6 and neighbor discovery over ATM. IETF draft-ietf-ion-ipv6nd-00.txt, 06/13/1996.

[6] Grenville J. Armitage. Multicast and multiprotocol support for ATM based internets. *ACM SIGCOMM Computer Communications Review*, 25(2):34–46, Apr 1995.

[7] Grenville J. Armitage. RFC2022: Support for multicast over UNI 3.0/3.1 based ATM networks, Nov 1996.

[8] Grenville J. Armitage and K. M. Adams. Using the common LAN to introduce ATM connectivity. In *Proceedings IEEE Computer Society 18th Conference on Local Computer Networks, Minneapolis, MN*. IEEE, Sep 1993.

[9] Grenville J. Armitage, Peter Schulter, Markus Jork, and Geraldine Harter. Transient neighbors for IPv6 over ATM. IETF draft-ietf-ion-tn-00.txt, 03/20/1997.

[10] Ran Atkinson. Security architecture for the Internet Protocol. IETF draft-ietf-ipsec-arch-sec-01.txt, 11/12/1996.

[11] Ran Atkinson, D. Haskin, and James V. Luciani. IPv6 over NBMA networks. IETF draft-ietf-ion-ipv6-nbma-01.txt, 11/14/1996.

[12] Fred Baker and Yakov Rekhter. Use of flow label for tag switching. IETF draft-baker-flow-label-00.txt, 11/1996.

[13] L. Berc, W. Fenner, R. Frederick, and S. McCanne. RFC2035: RTP payload format for JPEG-compressed video. Audio-Video Working Group, Oct 1996.

[14] Lou Berger. RSVP over ATM implementation guidelines. IETF draft-ietf-issll-atm-imp-guide-00.txt, 03/25/1997.

[15] Steven Berson and Lou Berger. IP integrated services with RSVP over ATM. IETF draft-ietf-issll-atm-support-03.txt, 03/26/1996.

[16] Uyless Black. *ATM: Foundation for Broadband Networks*, chapter 9, pages 226–227. Prentice-Hall Inc., Englewood Cliffs, NJ, USA, 1995.

[17] M. Borden, E. Crawley, J. Krawczyk, Fred Baker, and Steven Berson. Issues for RSVP and integrated services over ATM. IETF draft-crawley-rsvp-over-atm-00.txt, 1996.

[18] Rich Borden. FORE Systems and General Instrument announce plans to jointly develop ATM-based cable modem network system. `<http://www.fore.com/html/press/archive/PR61_09.html>`, 1996.

[19] R. Braden. RFC1122: Requirements for internet hosts – communication layer, Oct 1989.

[20] R. Braden, D. Clark, and S. Shenker. Integrated services in the internet architecture: an overview. RFC1633: Integrated Service, Jul 1994.

[21] R. Braden, L. Zhang, Steven Berson, S. Herzog, and S. Jamin. Resource ReSerVation Protocol (RSVP) – version 1 functional specification. IETF draft-ietf-rsvp-spec-13.txt, 8/12/1996.

[22] Scott W. Brim. Cells in frames version 1.0: Specification, analysis, and discussion. Cornell University Web Page, `<http://cif.cornell.edu/specs/v1.0/CIF-baseline.html>`, Oct 1996.

[23] Caralyn Brown. Baseline text for MPOA. ATM Forum Technical Committee, ATMF/95-0824, Jun 1996.

[24] Ross Callon, Paul Doolan, Nancy Feldman, Andre Fredette, George Swallow, and Arun Viswanathan. A framework for MultiProtocol Label Switching. IETF draft-ietf-mpls-framework-00.txt, 05/12/1997.

[25] Ross Callon, Jason Jeffords, John Drake, Hal Sandick, and Joel M. Halpern. An overview of PNNI Augmented Routing. ATM Forum Contributions, ATMF/96-0354, Apr 1996.

[26] Ross Callon, Jason Jeffords, Hal Sandick, and Joel M. Halpern. Issues and approaches for Integrated PNNI. ATM Forum Contributions, ATMF/96-0355, Apr 1996.

[27] Ross Callon, George Swallow, Loa Andersson, Joel M. Halpern, Bryan Gleeson, Jason Jeffords, Allwyn Sequeira, John Drake, Andre Fredette, and Ed Sullivan. The relationship between MPOA and Integrated PNNI. ATM Forum Contributions, ATMF/96-0352, Apr 1996.

[28] Derya H. Cansever. NHRP protocol applicability statement. IETF draft-ietf-ion-nhrp-appl-01.txt, 03/1997.

[29] Steve Casner. Minutes of the audio/video transport working group. Newsletter of the BC Advanced Systems Institute, DAVI Technology Standards Watch Group, Feb 1997.

[30] Isidro Castineyra, J. Noel Chiappa, and Martha Steenstrup. RFC1992: The Nimrod Routing Architecture, Aug 1996.

[31] J. Noel Chiappa. RFC1753: IPng technical requirements of the Nimrod Routing and Addressing Architecture, Dec 1994.

[32] ATMF Committee. ATM User-Network Interface specifications, version 3.0. Technical Report af-uni-0010.001, ATM Forum, 1993.

[33] ATMF Committee. ATM User-Network Interface specifications, version 3.1. Technical Report af-uni-0010.002, ATM Forum, 1994.

[34] ATMF Committee. ATM User-Network Interface specifications, version 4.0. Technical Report 94-1018, ATM Forum, 1995.

[35] The ATM Forum Technical Committee. Interim Inter-Switch Signaling Protocol (IISP) specification v1.0. Technical Report af-pnni-0026.000, ATM Forum, Dec 1994.

[36] The ATM Forum Technical Committee. Private Network-Network Interface specification version 1.0 (PNNI 1.0). Technical Report af-pnni-0055.000, ATM Forum, Mar 1996.

[37] The ATM Forum Technical Committee. Private Network-Network Interface specification version 1.0 addendum (Soft PVC MIB). Technical Report af-pnni-0066.000, ATM Forum, 1996.

[38] The ATM Forum Technical Committee. Addendum to PNNI v1.0 for ABR parameter negotiation. Technical Report af-pnni-0075.000, ATM Forum, 1997.

[39] Ascend Communications. Ascend grf ip switch. Ascend White Paper, `<http://www.ascend.com/>`, Feb 1997.

[40] Eric S. Crawley. Integrated services over ATM LAN Emulation (LANE) networks. IETF draft-ietf-issll-lane-00.txt, 11/27/1996.

[41] Bruce Davie, Paul Doolan, J. Lawrence, K. McCloghrie, Yakov Rekhter, E. Rosen, and G. Swallow. Use of Tag Switching with ATM. IETF draft-davie-tag-switching-atm-00.txt, 11/1996.

[42] Fred Dawson. ATM to the home starts long slow climb in 1996. *Inter@ctive Week*, `<http://www.zdnet.com/intweek/print/960129/infra/doc2.html>`, Jan 1996.

[43] S. Deering and R. Hinden. RFC1883: Internet Protocol, version 6 (IPv6) specification, Dec 1995.

[44] Stephen Deering, Deborah L. Estrin, Dino Farinacci, Van Jacobson, Ching-Gung Liu, and Liming Wei. The PIM architecture for wide-area multicast routing. *IEEE/ACM Transactions on Networking*, 4(2):153–162, April 1996. PIM.

[45] Steven Deering, Deborah Estrin, Dino Farinacci, Mark Handley, Ahmed Helmy, Van Jacobson, Ching gung Liu, Puneet Sharma, David Thaler, and Liming Wei. Protocol Independent Multicast–Sparse Mode (PIM-SM): Motivation and architecture. IETF draft-ietf-idmr-pim-arch-04.txt, 10/24/1996.

[46] Steven Deering, Deborah Estrin, Dino Farinacci, and Van Jacobson. Protocol Independent Multicast version 2, Dense Mode specification. IETF draft-ietf-idmr-pim-dm-spec-05.txt, 05/28/1997.

[47] L. Delgrossi and L. Berger. RFC1819: Internet Stream Protocol version 2 (ST2) protocol specification - version st2+, Aug 1995.

[48] Paul Doolan, Bruce Davie, Dave Katz, Yakov Rekhter, and Eric Rosen. Tag Distribution Protocol. IETF draft-doolan-tdp-spec-01.txt, 05/23/1997.

[49] Brian Dorling, Daniel Freedman, Chris Metz, and Jaap Burger. *Internetworking Over ATM, An Introduction*. IBM Technical Support Organization. Prentice-Hall, Inc., Upper Saddle River, N.J., USA, 1996.

[50] Hiroshi Esaki, Yasuhiro Katsube, Shigeo Matsuzawa, Akiyoshi Mogi, Kenichi Nagami, and Tatsuya Jinmei. Cell switch router. Toshiba Corp. White Paper, `<ftp://ftp.wide.toshiba.co.jp/pub/csr/whpaper.doc>`, November 1996.

[51] Hiroshi Esaki, Yasuhiro Katsube, Kenichi Nagami, Paul Doolan, and Yakov Rekhter. IP address resolution and ATM signaling for MPLS over ATM SVC services. IETF draft-katsube-mpls-over-svc-00.txt, 06/01/1997.

[52] Deborah Estrin, Dino Farinacci, Ahmed Helmy, David Thaler, Steven Deering, Mark Handley, Van Jacobson, Ching-Gung Liu, Puneet Sharma, and Liming Wei. Protocol Independent Multicast–Sparse Mode (PIM–SM): Protocol specification. IETF draft-ietf-idmr-pim-sm-spec-10.ps, 03/15/1997.

[53] Dino Farinacci, D. Meyer, and Yakov Rekhter. Intra-LIS IP multicast among routers over ATM using sparse mode PIM. IETF draft-ietf-ion-intralis-multicast-00.txt, 04/25/1997.

[54] Nancy Feldman and Arun Viswanathan. ARIS specification. IETF draft-feldman-aris-spec-00.txt, 03/1997.

[55] Kenji Fujikawa. Another ATM signaling protocol for IP (IP-SVC). IETF draft-fujikawa-ipsvc-01.txt, 11/1996.

[56] Mark W. Garrett and Marty Borden. Interoperation of controlled-load and guaranteed services with ATM. IETF draft-ietf-issll-atm-mapping-02.txt, 03/26/1997.

[57] Tom R. Halfhill. Break the bandwidth barrier. *Byte*, pages 68–80, Sep 1996.

[58] Joel M. Halpern. Implementing MPOA in ATM edge devices. *Communication Systems Design*, 2(10):25–34, Oct 1996.

[59] Joel M. Halpern. the architecture and status of PNNI. Newbridge White Paper, `<http://www.vivid.newbridge.com/mpoa/documents/PNNI/index.html>`, 1997.

[60] Juha Heinanen. RFC1483: Multiprotocol encapsulation over ATM Adaptation Layer 5, Jul 1993.

[61] Tom R. Henderson. Design principles and performance analysis of SSCOP: A new ATM Adaptation Layer protocol. *ACM SIGCOMM Computer Communications Review*, 25(2):14, Apr 1995.

[62] Robert M. Hinden and Stephen E. Deering. IPv6 technical overview. In Scott O. Bradner and Allison Mankin, editors, *IPng: Internet Protocol Next Generation*, chapter 8, pages 217–232. Addison-Wesley Publishing Company, Reading, MA, USA, first edition, 1996.

[63] D. Hoffman, G. Fernanco, and V. Goyal. RFC2038: RTP payload format for MPEG1/MPEG2 video. Audio-Video Working Group, Oct 1996.

[64] Yasuhiro Katsube, Kenichi Nagami, and Hiroshi Esaki. Router extensions for ATM: Overview. IETF draft-rfced-info-katsube-00.txt, 11/1996.

[65] Tim Kwok. A vision for residential broadband services: ATM-to-the-home. *IEEE Network*, 9(5):14–28, Jul 1995.

[66] Mark Laubach. RFC1577: Classical IP and ARP over ATM, Jan 1994.

[67] Mark Laubach. RFC1754: IP over ATM working group's recommendations for the ATM forum's multiprotocol BOF version 1, Jan 1995.

[68] Mark Laubach and Joel Halpern. Classical IP and ARP over ATM. IETF draft-ietf-ion-classic2-02.txt, 03/26/1997.

[69] Will E. Leland, Murad S. Taqqu, Walter Willinger, and Daniel V. Wilson. Protocol architecture for multimedia applications over ATM networks. *ACM SIGCOMM Computer Communication Review*, 23(4):pages, October 1993.

[70] Arthur Lin, Bruce Davie, and Fred Baker. Tag Switching support for classes of service. IETF draft-lin-tags-cos-00.txt, 12/1996.

[71] James V. Luciani. Classical IP to NHRP transition. IETF draft-ietf-ion-transition-02.txt, 05/1997.

[72] James V. Luciani. A distributed NHRP service using SCSP. IETF draft-ietf-ion-scsp-nhrp-01.txt, 06/1977.

[73] James V. Luciani, Grenville J. Armitage, and Joel Halpern. Server Cache Synchronization Protocol (SCSP). IETF draft-ietf-ion-scsp-01.txt, 04/1997.

[74] James V. Luciani, Dave Katz, David Piscitello, and B. Cole. NBMA Next Hop Resolution Protocol (NHRP). IETF draft-ietf-rolc-nhrp-11.txt, 03/05/1997.

[75] Steven McCanne and Van Jacobson. vic - video conferencing tool *(LBL Website)*. <http://www-nrg.ee.lbl.gov/vat/>, 1997.

[76] Kenichi Nagami, Yasuhiro Katsube, Yasuro Shobatake, Akiyoshi Mogi, Shigeo Matsuzawa, Tatsuya Jinmei, and Hiroshi Esaki. RFC2129: Toshiba's Flow Attribute Notification Protocol (FANP) specification, Apr 1997.

[77] T. Narten, E. Nordmark, and W. Simpson. RFC1970: Neighbor Discovery for IP version 6 (IPv6), Aug 1996.

[78] Peter Newman, W. L. Edwards, R. Hinden, E. Hoffman, F. Liaw, T. Lyon, and G. Minshall. RFC1953: Ipsilon Flow Management Protocol specification for IPv4 version 1.0, May 1996.

[79] Peter Newman, W. L. Edwards, R. Hinden, E. Hoffman, F. Liaw, T. Lyon, and G. Minshall. RFC1954: Transmission of Flow Labelled IPv4 on ATM data links, Ipsilon version 1.0, May 1996.

[80] Peter Newman, W. L. Edwards, R. Hinden, E. Hoffman, F. Ching Liaw, Thomas Lyon, and Greg Minshall. RFC1987: Ipsilon general switch management protocol specification version 1.1, Aug 1996.

[81] Peter Newman, Greg Minshall, and Tom Lyon. IP Switching: ATM under IP. *IEEE/ACM Transactions on Networking*, expected 1997. <http://www.ipsilon.com/technology/papers/newman0001.pdf>.

[82] Peter Newman, Greg Minshall, Tom Lyon, and Larry Huston. IP Switching and gigabit routers. *IEEE Communications Magazine*, 35(1):64–69, Jan 1997. <http://www.ipsilon.com/technology/papers/ieee_comm96.pdf>.

[83] Yoshihiro Ohba, Hiroshi Esaki, and Yasuhiro Katsube. Comparison of Tag Switching and Cell Switch Router. IETF draft-ohba-tagsw-vs-csr-00.txt, 04/97.

[84] Masataka Ohta. Multicast inscalability over large cloud. IETF draft-ohta-mcast-large-cloud-01.txt, 6/1996.

[85] Raif O. Onvural, Hal Sandick, and Rao Cherukuri. Structure and use of signaling in B-ISDNs. *Computer Networks and ISDN Systems*, 28(3):307–323, Jan 1996.

[86] Craig Partridge. RFC1809: Using the flow label field in IPv6, Jun 1995.

[87] Guru Parulkar, Douglas C. Schmidt, and Jonathan S. Turner. IPatm: a strategy for integrating IP with ATM. In *ACM SIGCOMM Symposium on Communications Architectures and Protocols, Cambridge, MA*. ACM, Sep 1995.

[88] Jon Postel. RFC791: Internet Protocol – DARPA internet program protocol specification, Sep 1981.

[89] M. Rajagopal and S. Sergeant. Internet Stream Protocol version 2 (ST2) protocol state machines - version ST2+. IETF draft-ietf-st2-state-02.txt, 10/1996.

[90] Yakov Rekhter. NHRP for destinations off the NBMA subnetwork. IETF draft-ietf-ion-r2r-nhrp-00.txt, 02/1997.

[91] Yakov Rekhter, Bruce Davie, Dave Katz, Eric Rosen, and George Swallow. Tag switching architecture overview. IETF draft-rfced-info-rekhter-00.ps, 9/1996.

[92] Yakov Rekhter and Tony Li. RFC1887: An architecture for IPv6 unicast address allocation, Dec 1995.

[93] Eric C. Rosen, Yakov Rekhter, Daniel Tappan, Dino Farinacci, and Guy Fedorkow. Label Switching: Label stack encodings. IETF draft-rosen-tag-stack-01.txt, 03/1997.

[94] Dragos Ruiu. An overview of MPEG. In *1996 Digital Video Test Symposium*, volume 2, chapter 9, pages 21–33. Hewlett-Packard Company, first edition, 1996.

[95] Khalid Sayood. *Introduction to Data Compression*. Multimedia Information and Systems. Morgan Kaufmann Publishers, Inc., San Francisco, CA, USA, 1996. audio, video, data compression.

[96] P. Schulter, M. Jork, and G. Harter. A framework for IPv6 over ATM. IETF draft-ietf-ion-ipv6atm-framework-00.txt, 6/13/1996.

[97] H. Schulzrinne. Issues in designing a transport protocol for audio and video conferences and other multiparticipant real-time applications. IETF draft-ietf-avt-issues-01.ps, 10/20/1993.

[98] H. Schulzrinne. RFC1890: RTP profile for audio and video conferences with minimal control. IETF Audio-Video Transport Working Group, Jan 1996.

[99] H. Schulzrinne, S. Casner, R. Frederick, and V. Jacobson. RFC1889: RTP: A transport protocol for real-time applications. Audio-Video Working Group, Jan 1996.

[100] Joshua Seeger. BBN's Multi-Gigabit Router for High-Speed IP Internetworking. BBN Web Page, <http://www.net-tech.bbn.com/mgr/pp-mgr-ovw/sld001.htm>, Jun 1996.

[101] Chuck Semeria and Tom Maufer. Introduction to IP multicast routing. IETF draft-ietf-mboned-intro-multicast-02.txt, 03/20/1997.

[102] Mohammad Shabani, Zikrun Badri, and Gillermo Kretchmar. IBM 8260 as a campus ATM switch. Technical Report SG24-5003-00, IBM International Technical Support Organization, June 1996.

[103] Dorgham Sisalem, Henning Schulzrinne, and Christian Sieckmeyer. The network video terminal. In *HPDC Focus Workshop on Multimedia and Collaborative Environments (Fifth IEEE International Symposium on High Performance Distributed Computing), Syracuse, New York*. IEEE Computer Society, Aug 1996.

[104] Michael Smirnov. EARTH - EAsy IP multicast Routing THrough ATM clouds. IETF draft-smirnov-ion-earth-02.txt, 03/22/1997.

[105] Frank Solensky. IPv4 address lifetime expectations. In Scott O. Bradner and Allison Mankin, editors, *IPng: Internet Protocol Next Generation*, chapter 1, pages 17–23. Addison-Wesley Publishing Company, Reading, MA, USA, 1 edition, 1996.

[106] M. Speer and D. Hoffman. RFC2029: RTP payload format of Sun's CellB video encoding. Audio-Video Working Group, Oct 1996.

[107] M. Speer and S. McCanne. RTP usage with layered multimedia streams. IETF draft-speer-avt-layered-video-01.ps, 6/10/1996.

[108] Burkhard Stiller. A survey of UNI signaling systems and protocols for ATM networks. *ACM SIGCOMM Computer Communications Review*, 25(2):13, Apr 1995.

[109] Muneyoshi Suzuki. ST2+ over ATM protocol specification - UNI 3.1 version. IETF draft-suzuki-st2-over-atm-01.txt, 03/1997.

[110] Rajesh Talpade and Mostafa H. Ammar. RFC2149: Multicast server architectures for MARS-based ATM multicasting, May 1997.

[111] S. Thomson and T. Narten. RFC1971: IPv6 stateless address autoconfiguration, Aug 1996.

[112] T. Turletti and Christian Huitema. RFC2032: RTP payload format for H.261 video streams. Audio-Video Working Group, Oct 1996.

[113] unknown. Video over ATM and existing networks. Cisco White Paper, <http://www.cisco.com/warp/public/730/General/vatm_wp.pdf>, Nov 1995.

[114] Arun Viswanathan, Nancy Feldman, Rick Boivie, and Rich Woundy. ARIS: Aggregate Route–Based IP Switching. IETF draft-viswanathan-aris-overview-00.txt, 03/1997.

[115] Ric Waller. Low Level ATM Interface. Technical report, IBM Networking Systems Architecture, <waller@vnet.ibm.com>, May 1995.

[116] D. Williams, R. Guerin, and D. Kandlur. ATM virtual circuit identification support for an RSVP-based service. IETF draft-williams-issll-vcuse-00.txt, 9/17/1996.

[117] Loring Wirbel. OEMs skirt ATM spec freeze by backing cornell u. project. <http://www.techweb.com/se/directlink.cgi?EET19960422S0044.html>, Jun 1996.

[118] R. Woundy, A. Viswanathan, N. Feldman, and Rick Boivie. ARIS: Aggregate Route-based IP Switching. IETF draft-woundy-aris-ipswitching-00.txt, 11/1996.

[119] Lixia Zhang, Stephen Deering, Deborah Estrin, Scott Shankar, and Daniel Zappala. RSVP: A new Resource ReSerVation Protocol. *IEEE Network*, 7(5):8–18, Sep 1993.