Multi-protocol over ATM
the usability of MPOA for PTT Telecom

de Gier, M.E.

Award date:
1997

Technische Universiteit Eindhoven

Faculty of Electrical Engineering
Section of Information and Communication Systems

Master's Thesis:

Multi-protocol over ATM

The usability of MPOA for PTT Telecom

ing. M.E. de Gier

Coach	: ir. G.P. Buitenhuis (KPN Research Leidschendam)
Supervisor	: Prof.ir. F. van den Dool
Date	: May 1997

Author: ing. m.e. de Gier

Date: may 1997

Multi-protocol over ATM

The usability of MPOA for PTT Telecom

For internal use only at KPN and Eindhoven University of technology R&D-SV-97-306

KPN Research

Information sheet issued with Report R&D-SV-97-306

Title:	MULTI-PROTOCOL OVER ATM - The functionality of MPOA for PTT Telecom
Abstract:	This report gives an overview of the working principle of Multi-Protocol over ATM. A relationship of MPOA with Clip, Lane, and IP Switching is presented, concluding in some examples demonstrating when to use these techniques. This indicates the usability for PTT Telecom
Author:	Ing. M.E. de Gier
Reviewers:	Ing. C.I. de Gier
Department:	Communication Architectures and Open Systems (CAS)
Project:	ATM Corporate Networks
Supervisor(s):	prof. F. van den Dool at Eindhoven University of technology, ir. G.P. Buitenhuis at KPN Research Leidschendam
Date:	may 1997
	For internal use only at KPN and Eindhoven University of technology
Person responsible at KPN Research:	ir. G.P. Buitenhuis
Key words:	MPOA, ATM, LANE, NHRP, Classical IP over ATM, IP Switching, RSVP
Mailing list:	A.J.van der Bas, M.M.G.N.van den Bergh, M.Bolle (5X), G.P.Buitenhuis, W.van Essenberg, F. Hakimzadeh, D.de Jong, B.Kastelein, F.H.Klok, W.P.J.Kuling, J.H.Laarhuis, J.H. Muijnck, R. Schellius, G.E.Teseling, R.A.Veldhuijsen, F.E.W. Vervuurt, D.Wapstra, M.J.van de Weg A.M.A.Wouters and author (10x)

Preface

This report is the result of my final Master's thesis, performed as part of my study of Information Technology at the Eindhoven University of Technology. Between september 1996 and may 1997 I worked at the Dr. Neher Laboratory of KPN research in Leidschendam at the department Communication Architectures and Open Systems (CAS).

The basis of my thesis is the "Multi-Protocol over ATM" from the ATM Forum. To perform my thesis I've been able to give my own contribution. Initially, I collected a lot of information. After reading all this information, I concluded that it will be difficult to outline the thesis. During my time at KPN research several new technologies are presented by several companies, like Ipsilons "IP switching" and Cisco "Tag switching" and Gigabit ethernet. To come to a conclusion I have framed my idea's, to indicate where I've focused my attention. Finalising this thesis gives me a great satisfaction and increases my interest in the telecommunication market.

At this point, I would like to take the opportunity to thank Prof. Ir. F. van den Dool for taking the responsibility for my graduation project, his interest in my progress and his useful comments. Also I like to thank my coach Trude Buitenhuis for the time she has created to support me with my project and for the useful contributions and discussions. I also wants to thank Jan Gerard Snip and Marc van den Bergh for their contribution in the practical part of my project, doing research in LAN emulation and IP over ATM. I would like to thank my colleague students at KPN research at room LE135x for the pleasant time I have had at KPN research. There discussions about everything and nothing at our own coffeecorner where very relaxing, so I have had pleasant time at KPN research. Also I want to thank my little brother for reviewing my report and giving useful comments. Finally I thank my girlfriend for encouraging me to keep on doing my homework and for supporting my in difficult times.

May 1997,
Menno de Gier.

Table of contents

Management Summary

PROBLEM : Current Corporate Networks are beginning to get limited in their capability to keep up with the increasing user requirements. More capacity and flexibility is required within the network in order to support both traditional and new applications, like multimedia applications. ATM is generally seen as one of the network technologies able to fulfil this task. When introducing ATM into a corporate network as smoothly as possible, it is essential that ATM can work together with legacy LAN equipment. It is important that computers with legacy LAN interface cards are able to communicate among themselves over an ATM network and that communication is enabled by workstations with ATM adapter cards. Furthermore ATM workstations should be able to communicate. This can be accomplished using one of the ATM-based interconnection protocols Classical IP over ATM (Clip), LAN emulation (Lane), Multi-protocol over ATM (Mpoa) or IP Switching.

AIMS : The primary aim of this project is to give a clear overview of the current position of Multi-Protocol over ATM, abbreviated Mpoa. To realise this a comprehensive working principle is discussed and a comparison is made between Mpoa and Clip, Lane, and IP switching.

RELEVANCE FOR KPN : Mpoa can be employed as a migration of Lane to a world with both legacy LANs and ATM hosts. KPN is interested in having an outline of where and how to use this.

CONCLUSION : The different solutions, discussed in this document, of LAN/WAN networks are Clip, Lane, Mpoa, and IP switching. Each of these techniques has it own strong and weak points. Depending on these, the following conclusion, regarding to the point of usage, can be made:
Clip is best implemented as a backbone or in a network with a small amount of clients and without any bridges and hereby connecting and delivering fast network connection with a minimum complexity.
Lane is best implemented in a LAN network. It connects the various components like bridges and ATM attached hosts. Lane offers manageability and a virtual network, where clients can join a *elan* independent of the physical location in a WAN.
Clip and Lane have one major problem, they require routers to connect different subnets or elans together. Because routers can be a bottleneck **Mpoa** or **IP switching** is needed. Which of the two solutions is the best depends on the current situation. When Lane is already implemented, Mpoa is the mostly logical step. A second decision point depends on what the network manager wants. If he wants a complete and easily manageable network, he would most likely choose Mpoa. If other constraints are an issue, like budget, he might have to choose for IP switching. A negative aspect of Mpoa can be the duality in calculation of routing paths. In Mpoa a path has to be calculated on two ways, i.e. by the classical Layer 3 routing protocols and by the ATM signalling. IP switching does not have this negative aspect.

FOLLOW UP : To give a clear advise, Mpoa and IP switching should be investigated in more detail. One should perform several field tests on the support of QoS and overall performance. With this field tests special attention must be paid to the point where a flow is detected on the default path. This causes the set up of a shortcut. The above mentioned negative aspect of the duality in path calculations can be avoided by the use of the I-PNNI protocol, but more research has to be done on this subject. Because this thesis discusses the first version of Mpoa, which at the time of writing this is still a straw ballot version, more attention must be paid to the multicast Server and Quality-of-service possibilities of MPOA v2 when released.

Abbreviations

ATM

AAL : ATM Adaptation Layer
ABR : Available Bit Rate
ATM : Asynchronous Transfer Mode
CBR : Constant Bit Rate
CLIP : Classical IP over ATM
NNI : Network-Node Interface
QoS : Quality-of-Service
SAR : Segmentation and Reassemble sublayer
SVC : Switched Virtual Connection
UBR : Unspecified Bit Rate
UNI : User-Network Interface
VBR : Variable Bit Rate
VC : Virtual Channel
VCC : Virtual Channel Connection
VCI : Virtual Channel Identifier
VP : Virtual Path
VPC : Virtual Path Connection
VPI : Virtual Path Identifier

Lane

ELAN : Emulated LAN
BUS : Broadcast and Unknown Server
LANE : LAN emulation
LEC : LAN emulation Client
LECS : LAN emulation Configuration Server
LES : LAN emulation Sever
TLV : Type/Length/Value

Mpoa

IA : Internetwork layer Address
MARS : Multicast Address Resolution Server
MPC : MPOA Client
MPOA : Multi-Protocol over ATM
MPS : MPOA Server
NHRP : Next Hop Resolution Protocol
NHS : Next Hop Server
SCSP : Server Cache Synchronisation Protocol

Miscellaneous

ARP : Address Resolution Protocol
BGP : Border Gateway Protocol
CSMA/CD : Carrier Sense Multiple Access with Collision Detection
IEEE : Institute of Electronic and
IETF : Internet Engineering Task
IP : Internet Protocol
I-PNNI : Integrated Private Network Node Interface
IPX : InterPacket Exchange
IS-IS : Intermediate System - Intermediate System
ISO : International Standards
LAN : Local Area Network
LLC : Logical Access Control
LIS : Logical IP Subnet
MAC : Medium Access Control
OSPF : Open Shortest Path First
OSI : Open Systems
PNNI : Private Network-Network
RARP : Reverse Address Resolution
RIP : Routing Information Protocol
TCP : Transmission Control
UDP : User Data Protocol

1. INTRODUCTION

Current Corporate Networks are beginning to get limited in their capability to keep up with the increasing user requirements. More capacity and flexibility is required within the network in order to support both traditional and new applications, like multimedia applications. Networks are getting bigger and bigger, producing long route-calculation times, and a router-bottleneck is created. ATM is generally seen as one of the network technologies able to fulfil these tasks. For a smooth introduction of ATM into a corporate networks it is essential that ATM can work together with legacy LAN equipment. It is important that computers with legacy LAN interface cards are able to communicate among themselves over an ATM network and that communication is enabled by workstations with ATM adapter cards. Furthermore ATM workstations should be able to communicate. This can be accomplished using one of the ATM-based interconnection protocols Classical IP over ATM, LAN emulation, Multi-protocol over ATM.

1.1 PROBLEM DEFINITION AND PURPOSE

ATM has a lot of advantages like high speed, scalebility, the power to create virtual LANs, and the ability to support quality-of-service. All of this is positive, but not enough. There has to be a mechanism to integrate ATM into existing networks without having to replace existing networks such as Ethernet, token ring, and TCP/IP, IPX, AppleTalk infrastructures. In January 1994 the Internet Engineering Task Force (IETF) has defined a specification to provide native support of classical IP over ATM, abbreviated Clip. This means that networks with IP protocols (not other protocols, like IPX and AppleTalk) can run over ATM. With the same goal of supporting existing network traffic over ATM without modification, LAN emulation, abbreviated Lane, is developed by the ATM-Forum in January 1995. Both specifications, Clip and Lane are partial solutions to the problem to integrate ATM, so there are yet some problems i.e. currently none of the aforementioned specifications define how to leverage the Quality-of-Service capabilities of ATM networks. This is the point where Multi-protocol over ATM (Mpoa) breaks in. Mpoa expands Clip and Lane. In a nutshell, Mpoa does three things. In the first place, it defines a high performance, low latency way to route IP and other protocols across an ATM network. In the second place it enables network managers to build virtual subnetworks that span routed boundaries, so users can be grouped together regardless of there physical location. Finally, Mpoa permits applications to use ATMs Quality-of-Service capabilities. The question this thesis tries to answer is how Mpoa will accomplish these three things. To come to an answer, the functionality of Mpoa has to be

sorted out in more detail. The final goal of this is to provide information about the interest for PTT Telecom of Mpoa. Hereby the attention is focused on the following questions :

- "How does Mpoa work?",
- "What is the relationship with other protocols or solutions?",
- "How to decide when to use Mpoa and when to use other solutions?",
- "When is Mpoa of interest?", and
- "What does an implementation guideline look like?".

1.2 REPORT OUTLINE

This thesis is composed of 8 chapters, a list of abbreviations and a reference list. Each chapter discusses a particular item related to Multi-Protocol over ATM. Chapter one describes what the reason is why Mpoa is developed. Chapter two deals with the specification of ATM. ATM is the medium which Mpoa uses to transport data. This chapter deals with the different layers, physical, ATM and AAL layers. It also describes how a connection is established between different hosts in a ATM network and some practical issues of ATM are discussed. Chapter three deals with the different protocols used by, or instead of, Mpoa. After an introduction, the ISO-OSI model which provides a common basis for the development of networks is described. This basis is required to assure that different systems can be connected to each other. Chapter three ends with an overview/description of the Classical IP over ATM approach, LAN emulation, Next Hop Resolution Protocol, Routing Protocols including I-PNNI, Multicast Address Resolution Server, and several new technologies. Chapter four explains Mpoa, starting with a description of Mpoa, hereby dealing with the different components of Mpoa. Then a more detailed specification of Mpoa is given together with some other aspects of Mpoa, like the use of NHRP and Lane. Chapter four finalizes with an example of data transport in an Mpoa environment. When the distinguished features of Mpoa are known it is important to understand what the performance of Mpoa is in comparison with other protocols and solutions. This is treated in chapter five. Chapter five is divided into several subchapters. These subchapters describe the relationship of Mpoa with Clip, Lane, and IP Switching on several issues such as the number of network connections, geographical coverage, usage of the Resource reSerVation Protocol (RSVP) and a analysis of strong and weak features. When developing network-based applications it is important to know when to use Mpoa and in what situation it is recommended to use another protocol. This is explained in chapter six. This thesis finalises with conclusions and recommendations on Multi-Protocol over ATM and an overview of the references used to create this document.

2. ATM

ATM - Asynchronous Transfer Mode - is a specific packet-oriented way of exchanging information, that uses a time-division multiplex technique. The multiplexed information flow consist of blocks with fixed size, called cells.

2.1 INTRODUCTION

The Asynchronous Transfer Mode (ATM) [ATM] is a transmission technique that offers a unified approach to data transport. The term "*transfer*" comprises both transmission and switching aspects, so a transfer mode is a specific way of transmitting and switching information in a network. The term "*asynchronous*" refers to the fact that there's no need of a fixed timing relationship between the information rates used in the network and the information rate of the user. The user can submit data to the network at any time and at any rate. In ATM, all information to be transferred is packed into fixed-size slots called cells. These cells have a 48 octet information field and a 5 octet header. Whereas the information field is available for the user, the header field carries information that pertains to the ATM layer functionality itself, mainly the identification of a cell by means of a label, so that a cell can be routed through the network.
ATM is connection oriented. This means that there has to be a connection established before real data can be transported. A connection has two hierarchical levels, a virtual path level and a virtual channel level. A virtual path is a collection of virtual channels having the same virtual path identifier (VPI). Each connection in a ATM environment is identified with this

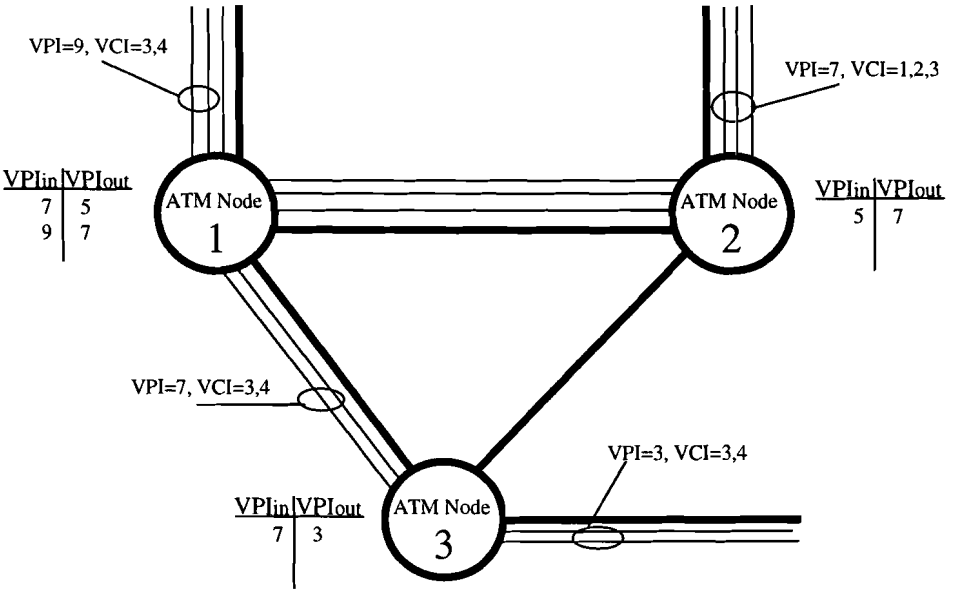


Figure 1: VCI/VPI switching

VPI. A virtual channel defines a connection between two ATM nodes (a switch or an end-station) and is identified be a virtual channel identifier (VCI). The VPI and VCI values are used in the header to identify a connection uniquely. Like the OSI model (see section 3.1), ATM has its own reference model which is defined in ITU-Recommendation I.321, see Figure 2. It consist of three layers, the physical layer, ATM layer, and ATM adaptation layer, and the layers above it.

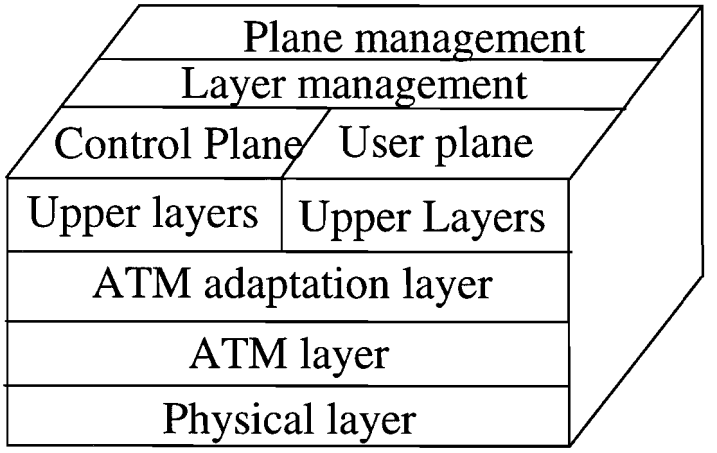


Figure 2 : ATM Reference Model

The user plane provides support for transfer of user information, associated controls, verification, and retransmission. The control plane handles the call control and connection control information and deals with the signalling flow necessary for initiating calls and connections, maintaining a call or the connection characteristics, ending and releasing calls. This part is called signalling and is discussed in section 2.6. Plane management functions satisfy the need for communication and they provide the co-ordination between the planes. Layer management performs functions relating to resources and parameters residing in protocol entities of the system. It handles Operation and Management Information flows that are specific for every layer. The following three sections discuss the physical layer, the ATM layer and the ATM adaptation layer in detail.

2.2 THE PHYSICAL LAYER

The physical layer deals with the physical medium : voltage, bit timing, and various other issues. ATM does not describe a particular set of rule, but instead describes that ATM cells may be sent on a wire or fiber by themselves, but they may also be packaged inside the payload of other carrier systems. In other words, ATM has been designed to be independent of the transmission medium. The main function of the physical layer is the transfer of cells from the ATM layer at the sender to the ATM layer of the receiver.

2.3 THE ATM LAYER

The ATM layer is independent of the physical medium used to transport the ATM cells and thus of the physical layer. The following main functions are performed by this layer:

- Multiplexing and demultiplexing of cells of different connection, based on there VPI, VCI values, into a single cell stream on a physical layer.
- A translation of the cell identification, which is required in most cases when switching a cell from one physical link to another, in an ATM switch or cross-connect. This

translation can be performed either on the VPI or VCI separately or on both simultaneously.
- Providing the user of a VCC or VPC with one Quality-of-Service class, out of a number of classes supported by the network.
- Management functions. The header of user information cells provides for a congestion indication and an ATM user to ATM user indication.
- Extraction/addition of the cell header before/after the cell is being delivered to/from the adaptation layer.
- Implementation of a flow control mechanism on the user-network interface.

In ITU-Recommendation I.361 the coding of ATM cells is described in detail. This recommendation specifically focuses on the cell structure and the ATM cell coding, and the ATM protocol procedures. As mentioned before, a cell contains a 48 octet information field and a 5 octet header. There are different cell headers at the UNI and NNI (see section 2.6 where the difference between the UNI, NNI is explained).

2.4 ATM ADAPTATION LAYER

The ATM adaptation Layer (AAL) enhances the service provided by the ATM layer to support the functions required by the higher layers. The functional description of the AAL is defined in ITU-Recommendation I.362, the specifications of the AAL are defined in ITU-Recommendation I.363. The basic principles of the AAL are the isolation of the higher layers from specific characteristics of the ATM layer by mapping the higher layer protocol data units (PDUs) into the information field of the ATM cell and visa-versa. To support services above the AAL, some interdependent functions must be performed in the AAL. These functions are organised in two logical sublayers, the Segmentation And Reassemble sublayer (SAR), and the Convergence Sublayer (CS). The prime functions of the SAR are (1) segmentation of higher layer information into a size suitable for the information field of an ATM cell and (2) reassemble of the contents of ATM-cell information fields into higher layer information. For example, a IP-packet of size 65535 bytes must be split-up into 65535/48=1366 ATM-cells. The prime functions of the CS is to provide the AAL services. The AAL consists of a variety of services like multiplexing, cell loss detection and timing recovery. There are four classes of services defined in ITU-Recommendation I.362, The classification is made with respect to the following parameters, timing relation, bit rate, and connection mode.

	Class A	Class B	Class C	Class D
Timing relation between source and destination	required		not required	
Bit rate	constant	Variable		
Connection mode	connection oriented			Connectionless

There is a relation between the classes for AAL and the AAL protocols, but its not a one-to-one relation. For every class of service there exists a certain combination of functions inside the AAL and these combinations are mapped onto AAL types. ITU-Recommendation I.363 describes a number of AAL types. Each type consists of a specific SAR sublayer and CS. Up to now four AAL types have been defined. The provision of constant bit rate services (CBR services) utilises AAL type 1 and is direct related to class A. AAL type 2 is related to class B and AAL type 3/4 is direct related to class C and class D. The last AAL is type 5 and is defined in the addendum no.1 of ITU-Recommendation I.363. AAL type 5 will be applied to variable bit-rate sources without timing relation between source and destination.

2.5 QUALITY OF SERVICE

Quality of service is an important issue for ATM networks, partly because they are used for real-time traffic, such as audio and video. When a virtual circuit is established by means of signalling, there has to be a contract defining the service. The contract between the customer and the network consists of three parts, the traffic to be offered, the service agreed upon, and the compliance requirements. The first part of the contract is the traffic descriptor and characterises of the load to be offered. The second part specifics the quality of service desired by the customer and accepted by the carrier. To have concrete traffic contracts, the ATM standard defines a number of QoS parameters. These parameters determine the speed of the traffic (*Peak Cell Rate, Sustained Cell Rate, and Minimum Cell Rate*), specification of the network characteristics (*Cell Error Ratio, Severely-Errored Cell Block Ratio, and Cell Missinsertion Rate*), the characteristics of the network measured at the receiver (*Cell Loss Ratio, Cell Transfer Delay, and Cell Delay Variation*), and the variation present in cell transmission times (*Cell Variation Delay Tolerance*). The third part of the contract tells what constitutes obey these rules. Note that the user does not explicit choose for any QoS parameter, but implicit selects a QoS class with this information[1]. Table 1 in Appendix A specifies the relationship between the QoS classes.

2.6 SIGNALLING

Signalling provides the procedures for dynamically establishing, maintaining and clearing ATM connections. The procedures are defined in terms of messages and the information elements used to characterise the ATM connection. ATM connections can be based on a permanent basis and one on switched basis. The Permanent Virtual Connection (PVC) is a connection that resists permanent and is always available to be reused at any time, like leased lines. The Switched Virtual Connections (SVC) have to be established each time they are used, like making a phone call. Signalling takes place in an separate AAL, called the Signalling AAL (SAAL). The Signalling AAL sends the signalling packets through a predefined VPC/VVC pair, namely VPC=0, VCI=5. In a SVC ATM network there are two instances where signalling takes place i.e. between a user and a network node and between two network nodes. The first is called the User-Network Interface (UNI) and the second Network-Network Interface (NNI). For both interfaces different protocols are defined. Between the user and network UNI-3.1, UNI-4.0 has recently been released, and between two network nodes PNNI or I-PNNI is defined or in development. Establishing a connection can be done using four message types, the *SETUP*, *CONNECT*, *CALL_PROCEEDING*, and *CONNECT_ACKNOWLEDGE*. The normal procedure for establishing a connection is for a host to send a *SETUP* message on the predefined virtual connection. The network responds with *CALL_PROCEEDING* to acknowledge receipt of the request. As the *SETUP* message propagates towards the destination, it is acknowledged at each hop by a *CALL_PROCEEDING*. When the *SETUP*

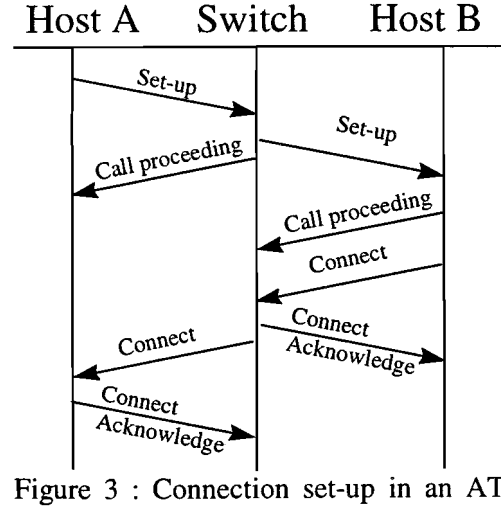


Figure 3 : Connection set-up in an ATM network.

[1] The next versions of the standards are more flexible; i.e. one can specify the QoS by every parameter.

arrives, the destination host response with a *CONNECT* to accept the call. The networks send a *CONNECT_ACKNOWLEDGE* message to indicate that it has received the *CONNECT* message. As the a *CONNECT* propagates back towards the originator, each switch receiving the *CONNECT* replies it with a *CONNECT_ACKNOWLEDGE*. This sequence is shown in Figure 3. Before a connection is set-up, according to the above mentioned way, the route through the network has to be calculated. This is done by a routing protocol. Several routing protocols are discussed in section 3.5.

2.7 STATUS OF ATM

At this moment there aren't much ATM networks yet. And if there are ATM solutions they are mainly used as a high-bandwidth back-bone. An ATM network in the near future will have a combination of legacy hosts and ATM attached hosts. These ATM attached hosts are using a legacy protocol with an ATM card inside to communicated with the ATM network. Many different vendors have ATM cards and ATM switches. Nowadays, many different solutions aren't working smoothly with each other. For example, a situation at KPN research, 3COM has an ATM-switch working with LAN emulation and the co-operation with an FORE ATM-switch has its difficulties, such as the resolving the location of the LANE services, and the fact that FORE uses a FORE specific LAN Emulation version. Another ATM network in the Netherlands is used by the royal land forces. They are using ATM as backbone, to create one high-bitrate-network, hereby using the LAN Emulation technique. In comparison with legacy LANs ATM has many advantages and disadvantages. One of the negative aspects of ATM is the overhead of the ATM cells. For every 48 octets there has to be a 5 octets header, so approximately 10 percent is wasted bandwidth. The capability of ATM to integrate data, voice, and video information on a common communication network, while providing Quality-of-Service to individual connections, is one of ATMs major benefits. Other advantages of ATM are its high speed, its scaleability, and the power to create virtual LANs. An other advantage is the usage of a wider address-space, namely 40 bytes, in contrary to the 48 bit of MAC-addresses and 32 bits for IP-addresses[2]. One of the advantage mentioned above is ATMs high speed i.e. ATM can offer 25Mb/s, 34Mb/s, 155MB/s up to 622Mb/s depending on the interface type. While Ethernet offers 10Mb/s bandwidth, fast-Ethernet in contrary offers 100Mb/s.

[2] IP version 4

3. Network protocols

This chapter discusses the protocols that are used at the moment and the protocols that can be used in the future. The discussed protocols are used by Mpoa or are competing Mpoa.

3.1 Introduction

As LANs became more and more popular, attempts were made to connect one network to another. These attempts were unsuccessful because each company create their own network and use their own standards of communication over the network. To solve this problem the international Standards Organisation (ISO) proposed a seven Layer model that would provide a common basis for the development that would allow different systems to be connected. This model is called the Open Systems Interconnection (OSI) Model. The seven Layers of the ISO-OSI model are discussed in section 3.2. The most used network protocol is the Internet Protocol (IP) and is discussed separately in section 3.3. Several network managers today are investigating the benefits and challenges associated with migrating their networks to Asynchronous Transfer Mode (ATM). ATMs inherent capabilities such as gigabit-level speeds, multi-service integration, virtual network support, and easy scalebility make it an attractive alternative for network growth. But network planners raise their caution flags when they consider how ATM technology will interoperate with their installed base of Ethernet and Token Ring equipment, data networking protocols, and legacy applications. Part of the problem is that there are different protocol specifications for running LAN traffic over ATM, i.e. the IETFs Classical IP over ATM, and the ATM Forum's LAN emulation specifications. Classical IP over ATM shares the same basic goal as LAN emulation, connecting legacy LANs and supporting existing network applications without modification, but it is focused exclusively on allowing IP traffic to run over ATM networks. Both protocols IP over ATM and LAN emulation are treated, IP over ATM in section 3.4 and LAN emulation in section 3.5.

Like legacy networks there is also a need for routing in the ATM network beside data transport. Section 3.6 deals with both the routing in legacy networks and routing in ATM networks. To take care of routing in ATM networks the ATM Forum Technical PNNI Subcommittee has been working on defining a routing and signalling protocol for use between ATM-switching systems. The PNNI protocol is further expanded to I-PNNI, so it can also work in an IP environment.

Both LAN emulation and IP over ATM solve important portions of the problem of enabling existing protocols and applications to operate over ATM, referred as ATM *internetworking*. However they both are restricted in the case of data transfers within a emulated LAN or subnetwork boundaries. For communication outside its subnetwork they must use a conventional router, even if the target is directly attached to ATM. Communication via a router implies higher latency due to additional layer 3 routing. The Next Hop Resolution

Protocol (NHRP) is being developed by a working group of the IETF in order to address this problem. Using Next Hop Servers that interact with routing protocols to propagate queries, NHRP provides an extended address resolution protocol that permits queries between different subnetworks. Because of the use of NHRP in the MPOA protocol, NHRP is discussed in section 3.7.
The IETFs Multicast Address Resolution Server (MARS) rounds out the suite of protocols for ATM internetworking. MARS is used to resolve internetworking layer multicast and broadcast addresses to either a list of ATM addresses, or to the ATM address of a Multicast Server that is responsible for distributing the data to the appropriate end-stations. MARS is discussed in section 3.8. In de last section several new technologies are discussed, like IPv6, Gigabit Ethernet, IP switching and Tag switching.
Furthermore it is important to note here, that it is not the application itself but always the network that initiates the connection set-up. This means that an application can not request a special QoS. These is not only the case with IP switching, but also with CLIP, LAN emulation, MPOA and legacy networks. To solves this problem, the Resource reSerVation Protocol (RSVP) is defined, to bring QoS from application to the network.

3.2 OSI model

The OSI model is shown in Figure 1 on this page. This model is based on a proposal developed by the International Standards Organisation (ISO) as a first step towards international standardisation of the protocols used in the various layers. The model is called the ISO OSI Reference Model. The OSI model has seven layers. The function of each layer is to provide service to the layer above it. Each layer in the model will only communicated with the layer above it or the layer below it.

From top to bottom the ISO model exists of :

- The ***physical layer*** is concerned with transmission of raw bits over a communication channel. It deals with mechanical, electrical, procedural interfaces, and the physical transmission medium.
- The main task of the ***data link layer*** is to take raw transmission facility and transform it into data that appears free of detected transmission errors to the above laying network layer. The data link layer consist of two parts, the ***MAC-sublayer*** and the ***LLC-sublayer***. The MAC sublayer is the bottom part of the data link layer and takes care of the access to the transmission channel. The LLC sublayer task care of the error control and flow control. There are several standard for LANs produced by the IEEE for the physical and data link layer and are known as IEEE 802. The IEEE 802.1 gives an introduction to the set of standards and defines the interface primitives. The IEEE 802.2 standard describes the upper part of the data link layer, that uses the LLC protocol. Parts IEEE 802.3 through 802.5 describes the tree LAN standards, the CSMA/CD (Ethernet), token ring, and token bus standards. Each standard covers the physical layer and the MAC-sublayer. Every IEEE 802 protocol has its own frame format. So has the IEEE 802.3 CSMA/CD a maximum payload of 1500 bytes and IEEE 802.4 token ring a 8182 bytes payload and IEEE 802.5 token bus has no limit on the maximum payload. The

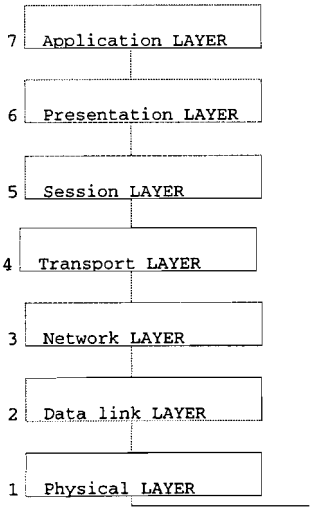


Figure 1: OSI Reference Model

MAC-sublayer of every protocol specifies a hardware address called the MAC-address. For each computer is this a unique address. This MAC-address is address is 6 bytes long. The idea is that any station can uniquely address any other station by just giving the right 6 byte number.

- The ***network layer*** is concerned with controlling the operation of the network. A part of this is the way packets are routed from source to destination. There are several protocols for this layer defined including IP, IPX, CLNP, Apple Talk and DECnet. A little bit more attention is paid to the most common protocol IP and to the popular protocol IPX in section 3.3.
- The basic function of the ***transport layer*** is to accept data from the session layer, split it into smaller units if needed, pass these to the network layer, and ensure that the all pieces arrive correctly at the receiving node. There are two several transport layer protocols, TCP, UDP, NCP, and SPX. The first two protocols, TCP and UDP, are both used in combination with the IP network protocol. And the other two, SPX and SP, are used with the IPX protocol.
- The ***session layer*** allows users on different machine to establish sessions between them. A session allows ordinary data transport, as does the transport layer, but it also provides enhanced services that are useful in some applications, such as synchronisation and token management.
- Unlike all the previous described layers, which are just concerned with moving bits reliably from here to there, the ***presentation layer*** is concerned with the syntacs and semantics of the information being transmitted.
- The ***application layer*** contains a variety of protocols that are commonly needed. Telnet, FTP, SMTP and DNS are protocols for the application layer.

3.3 Internet Protocol

One of the most used network layer protocols is IP (Internet Protocol), see [IP]. The IP routes data between hosts. IP is an example of a connectionless service. It permits the exchange of traffic between two hosts without any prior call set-up. An IP packet is made up of an IP header and a data unit. The IP header contains the source and destination IP address. Every host on a network with IP protocol has an IP-address. It consists of a network part and a host number. All IP addresses are 32 bits long and are written in dotted decimal notation, for example 194.171.110.38. In addition to IP, which is used for data transfer, there are several control protocols used in the network layer, including Internet Control Message Protocol (ICMP), Address Resolution Protocol (ARP), and Reverse Address Resolution Protocol (RARP). ICMP is used to test the network. One of the ICMP packets is *ECHO_REQUEST* (also known as *ping*) which is used to ask a machine whether or not it is alive. The ARP is used to resolve IP-addresses to MAC-addresses. This is necessary because IP-addresses cannot actually be used for sending packets since the data link layer hardware does not understand them. RARP solves the backwards problem i.e. resolves IP addresses corresponding to a given MAC-address. The characteristics of IP are referred to as a unreliable transport mechanism in the sense that :

1) No guarantee is given that the delivered datagram is error free.
2) No guarantee is given that the datagram is delivered at the destination, for example the datagrams may be lost.
3) No guarantee is given about the order in which datagrams are delivered, called *out-of-order.*
4) No guarantee is given about the delay or delay variation of the datagrams.
5) As a result of the previously mentioned lack of certainty, there is no guaranteed throughput.

Another popular network protocol is IPX (Internetwork Packet Exchange). IPX is similar to IP, except that it uses different address lengths. The supplier of IPX, Novell, has announced that it will change or remove its IPX-Protocol in favour of the TCP/IP protocol.

3.4 Classical IP over ATM

Classical IP over ATM defines a method of running IP over an ATM backbone. Classical IP over ATM, contained in the IETFs RFC 1577 document, was published in January 1994. The goal of the Classical IP over ATM specifications is to allow a compatible and interoperable implementation for transmitting IP packet over ATM using AAL5. The advantage of Classical IP over ATM is its simplicity. In a Permanent Virtual Connection (PVC) network all IP addresses are manually mapped to virtual connections. The user configures each station with a local address table that specifies the virtual connection corresponds to each IP-address on the ATM network. If a ATM network supports Switched Virtual Connections (SVCs), the end stations must have a way of mapping IP addresses into ATM addresses so connections can be set-up by the UNI. One additional protocol element is needed i.e. an ATM address resolution protocol (ATMARP) server. See figure 2 for the configuration of a classical IP over ATM environment. Classical IP over ATM is based upon two components :

- ***Logical IP Subnetwork (LIS)*** - A LIS allows users to be in the same IP subnetwork no matter where they are physically connected within the ATM network. Each LIS is a collection of IP end-stations. The concept considers only directly connected IP end-stations or routers. The following is a summary list of the requirements for a LIS configuration:
  - All members have the same IP network number. These are the first nine digits of the IP-address. As example 194.171.110 is the network part of IP-address 194.171.110.38.
  - For each LIS a single ATM group address has been configured that identifies all members of the LIS. Any packet transmitted with this address is delivered by ATM Service to all members of the LIS.
  - All stations within a LIS are accessed directly over ATM.
  - All stations outside of the LIS are accessed via a router.
- ***ATM Address Resolution Protocol (ATMARP) Server*** - The ATMARP server is a resource on each LIS that enables end-stations to resolve the ATM address of a given destination IP-address. The ATMARP server on each LIS automatically maintains a database for mapping IP-addresses to ATM-addresses.

In the LIS concept, each separate administrative entity configures its hosts within a closed logical IP subnetwork. Each LIS operates and communicates independently of other LISs over the same network providing ATM. Hosts connected to ATM communicate directly to other hosts within the same LIS. Communication to hosts outside of an individual LIS is provided via an IP router. This router can simply be a station attached to the ATM Service that has been configured to be a member of both logical IP subnetworks. This configuration results in a number of disjoint LISs operating over the same network supporting the ATM Service. It is recognised that with this configuration, hosts of differing IP networks would communicate via an intermediate router even though a direct path over the ATM Service may be possible.

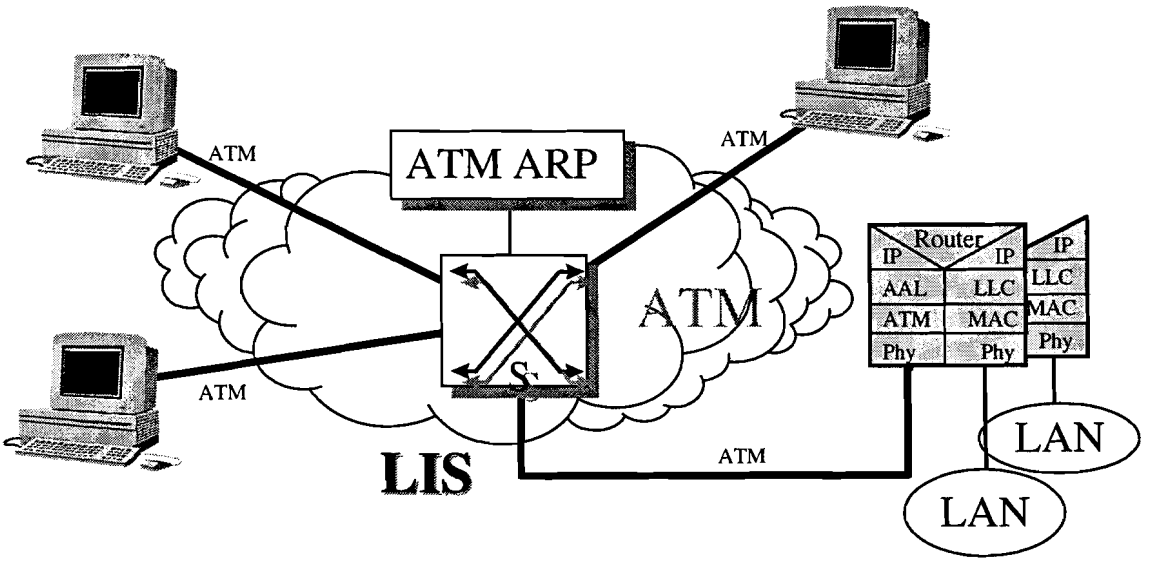


figure 2: Classical IP over ATM configuration (inclusive IP/ATM stack)

When a workstation needs to resolve an IP-address to an ATM-address, it contacts the ARP server that is responsible for all users within the LIS. The communication with a user on a different subnetwork takes place through a router connected to both subnetworks. For the multicast and broadcast addresses there is no solution in classical IP over ATM.

3.5 LAN emulation

The ATM forum has published the LAN emulation specification version 1.0 in January '95. The LAN emulation Version 1.0 specifies how the LAN emulation client (LEC) interacts with the LAN emulation service across the User-to-Network Interface (UNI). LAN emulation Version 2.0 is going to be published June '97 and adds a LANE User-to-Network Interface (LUNI), multiple servers/ busses, and multicast server to the specification. LAN emulation operates at the media access control (MAC) layer and enables legacy Ethernet, Token Ring, or FDDI (indirect) traffic to run over ATM without requiring modifications to applications, network operating systems, desktop adapters, or wiring and cabling. The LAN emulation specification is based on a client-server implementation model. A emulated LAN version 1.0consists of one LAN emulation Service and multiple Clients communicating through the UNI. LAN emulation components include the following:

- ***LAN emulation Client (LEC)*** - End systems that support LAN emulation require the implementation of a LEC. The LEC emulates an interface to a legacy LAN to the higher-level protocols. It performs data forwarding, address resolution, and registration of MAC addresses with the LANE server and communicates with other LECs via ATM virtual channel connections (VCCs).
- ***LAN emulation Server (LES)*** -The LES provides a central control point for all LECs. LECs maintain a 'Control Direct VCC' to the LES to forward registration and control information. The LES maintains a point-to-multipoint VCC, known as the 'Control Distribute VCC' to all LECs. The 'Control Distribute VCC' is used only to forward control information. As new LECs join the ATM emulated LAN, each LEC is added as a leaf to the control distribute tree.

- ***Broadcast and Unknown Server (BUS)*** -The BUS acts as a central point for distributing broadcasts and multicast. ATM is essentially a point-to-point and point-to-multipoint technology without "any-to-any" or "broadcast" support. LANE solves this problem by centralising the broadcast support in the BUS. Each LEC must set up a 'Multicast Send VCC' to the BUS. The BUS then adds the LEC as a leaf to its point-to-multipoint VCC, known as the 'Multicast Forward VCC'. The BUS also acts as a multicast server.
- ***LAN emulation Configuration Server (LECS)*** -This server maintains a database of LECs and the emulated LANs that they belong to. It accepts queries from LECs and responds with the appropriate emulated LAN identifier i.e. the ATM address of the LES that serves the appropriate emulated LAN. This database is maintained by the network administrator.

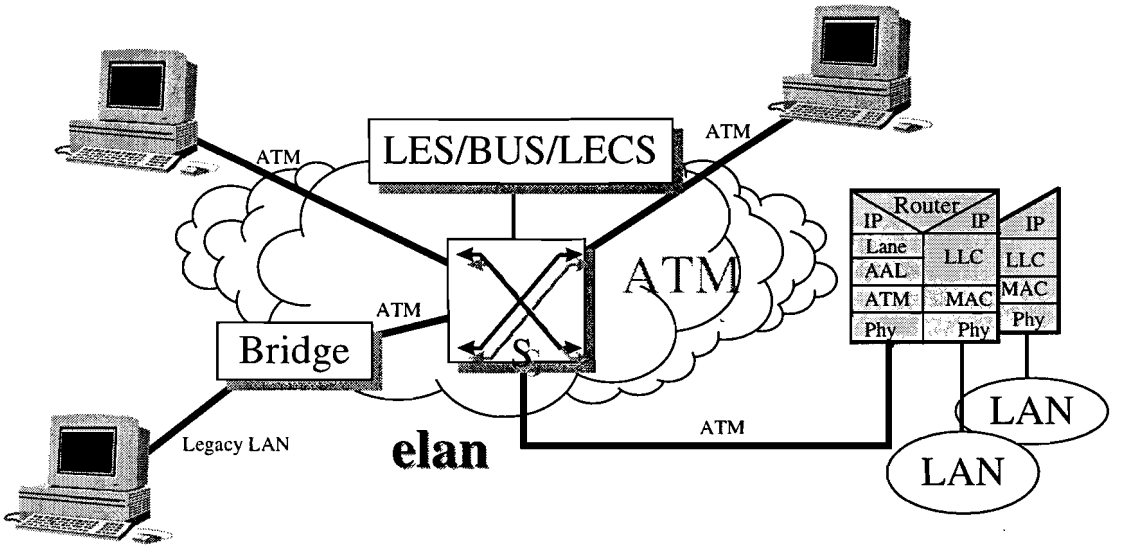


figure 3 : LAN emulation Configuration

Before any communication is possible a LEC must initialise itself. This includes joining a emulated LAN. The first step to join an emulated LAN is to find the LECS. This can be done in one of the following ways :

1. A LEC can send a ILMI message to the ATM switch to obtain the ATM address of the LECS.
2. The LEC can use a "well-known-address".
3. The LEC could use a predefined VCC to the LECS.
4. The LECS can be bypassed completely by configuring the ATM-address of a LES in the LEC.

Once a LEC locates the LECS, a connection, called 'Configuration Direct VCC', is set-up to the LECS. Next, the LEC must find the ATM-address of the LES. This is done by sending a *LE_CONFIGURE_REQUEST*-packet to the LECS over the 'Configuration Direct VCC'. The LECS answers this with a *LE_CONFIGURE_RESPONSE*-packet. This response contains the ATM-address of the LES belong to the emulated LAN the LEC wishes to join. See figure 4 for the different existing VCC between LEC and LAN emulation services. The second step is setting up a connection, called 'Control Direct VCC', to the LES. The LES on his turn, adds the LEC to a point-to-multipoint connection called 'Control Distribute VCC'. If this connection does not exists the LES creates it. After this the LEC sends, via the 'Control Distribute VCC', a *LE_JOIN_REQUEST*-packet to the LES. With this packet it registers its MAC-address and associated ATM-address. The LES answers with a *LE_JOIN_RESPONSE*-packet to indicated that the LEC has join a emulated LAN. The third step is to obtain the ATM-address of the BUS. This is done by sending a *LE_ARP_REQUEST*-packet, to the LES, via the 'Control Direct VCC'. The LES on his turn answers with the ATM-address of the BUS, by sending a

LE_ARP_RESPONSE-packet, via the 'Control Direct VCC'. With the BUS's ATM-address the LEC sets up a connection, called 'Multicast Send VCC', to the BUS. The BUS on his turn adds the LEC to a point-to-multipoint connection called 'Multicast Forward VCC'. If this connection does not exists the LES creates it. After this the initialisation phase is finished and the LEC is ready to handle data packets.

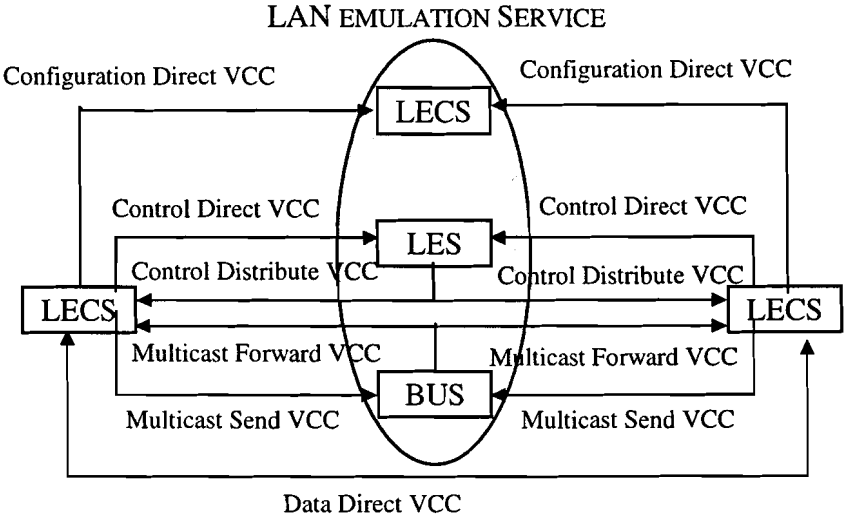


figure 4 : LAN emulation Connections

If two stations wants to exchange IP data-traffic, and both connected to the same emulated LAN, they must first, like normal IP traffic, translate an IP-address into a MAC-address. This is done by sending an ARP request (an IP-ARP of course) encapsulated by LAN emulation to the LEC at the other side. Because the MAC-address is not known yet,, it is unknown data, thus send to the BUS through the 'Multicast Send VCC'. The BUS forwards this packet to every LEC connected via the 'multicast Forward VCC'. The destination LEC resolves the IP-address and answers this ARP-request by sending a *LE_ARP_REPLY*-packet back to the BUS, through the 'Multicast Send VCC'. The BUS forwards the ARP-reply to the source LEC. Now that the MAC-address is known it will be translated by LEC to an ATM-address. This is done by sending a *LE_ARP_REQUEST*-packet so that it can be resolved. This packet is send to the LES, via the 'Control Direct VCC'. Because the destination LEC has also registered itself, the LES can answer the request with a *LE_ARP_RESPONSE*-packet containing the ATM-address. Now that the LEC knows every detail, the LEC can set-up a data connection, called 'Data direct VCC'. After the connection is set-up, LAN emulation sends a *READY_IND*-packet, to indicate that the ATM connection can be used for data-communication. If the destination LEC is located on a different emulated LAN, the above mention procedure is repeated twice. Once between the initiating LEC to a router and once between the router and the destination LEC.

3.6 Routing Protocols

In every network packets must be sent from source to destination. A routing protocol determines the path of the packet through the network. There are two levels of routing protocols, i.e. within a network, called "interior gateway protocols", and between networks, "exterior gateway protocols". One of the first routing protocols is an interior gateway protocol, the Routing Information Protocol (RIP). This protocol is a distance vector protocol. This means that each router is equipped with a table (vectors) giving the best known distance to each destination and indicating which line to use to get there. The tables are updated by

exchanging information with there neighbours. The successor of RIP is a link-state protocol called Open Shortest Path First (OSPF). A link state protocol consist of five parts. The first one is to detect its adjacent routers and network-addresses, by means of sending hello-packets. Secondly the cost to every neighbours is detected by measuring the round trip time. Thirds, a list is made with the above mentioned data. Sequentially, this data is sent to all the other routers. Finally the shortest path to every router is calculated. Another important link state protocol is IS-IS (Intermediate System-Intermediate System) and is designed for DECnet for use with its connectionless network layer protocol (CLNP). IS-IS is later modified to handle other protocols, including IP and IPX (a version called NLSP).
Between networks a different protocol, the Border Gateway Protocol (BGP), is used. The distinction between BGP and a interior gateway protocol is that policies need to be considered. Typical policies involve political, security, or economic considerations. From the point of view of BGP, the world consist of BGP routers and lines connecting them. BGP is fundamentally a distance vector protocol, but instead of the cost it keeps track of the exact path used.
Since ATM is connection oriented, a connection set-up must be done before data-packets are transmitted. Connection set-ups need to be routed from the requested node through the ATM network to the destination. In an ATM network a routing protocol enables communication between different ATM switches, or between a ATM switch and a private ATM switching system. It enables an ATM switch to find a path to any other switch. Within an ATM-

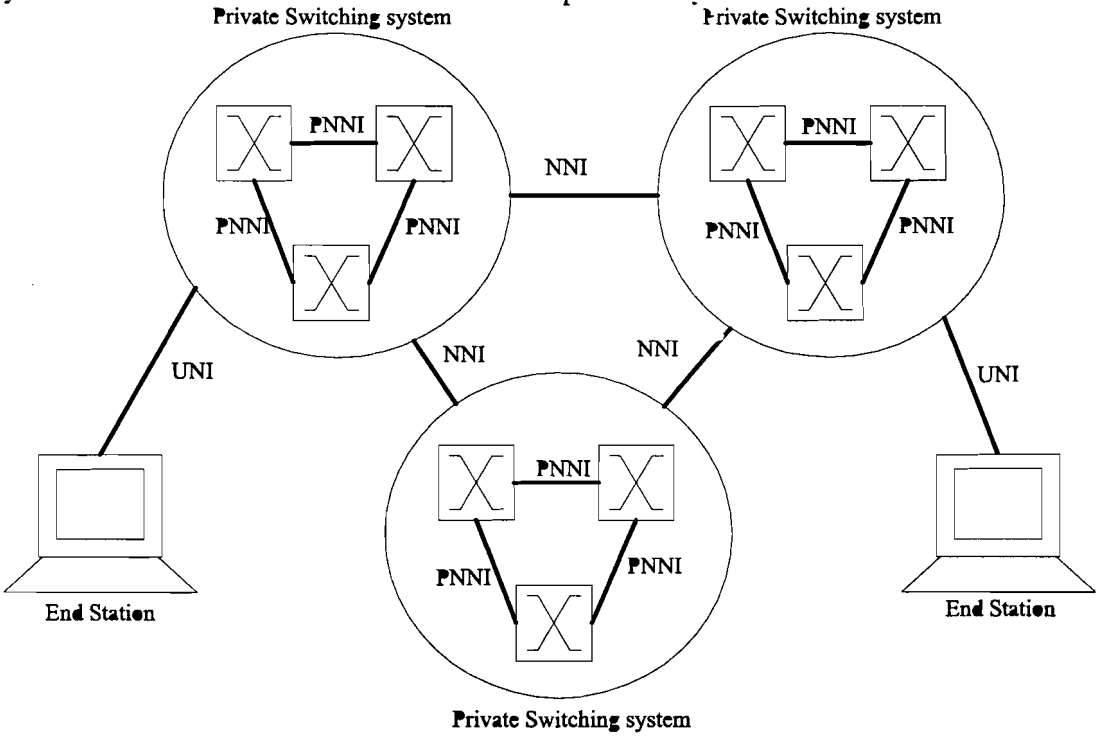


figure 5 : ATM network

network there are several different routing hierarchies, i.e. User-Network Interface (UNI), Network-Network Interface (NNI), and Private-NNI (PNNI) protocol. The UNI protocol connects end-stations to switches and a PNNI protocol connect switches within private ATM-networks. Finally, the NNI protocol connects the switching systems. **figure 5** shows the difference between them. The UNI and NNI protocols are routing protocols for a ATM network, what legacy routing protocols, like IS-IS, OSPF, RIP, and BGP are for the legacy networks. PNNI is a link state routing protocol and standardises switch-to-switch signalling and topology information distribution within the ATM-network. One of the features of PNNI is the support of Quality-of-Service.

PNNI allows the topology information within each switching system to be distributed to all attached switching systems. The organised topology information represents the hierarchical view of the ATM network based on the ATM address structure. Without the hierarchical nature of PNNI, every ATM switching system would have to maintain a complete picture of the total topology. Ultimately, the distribution of topology information becomes the mechanism used to efficiently compute end-to-end paths through the network. Like OSPF, PNNI is a hierarchical, link state routing protocol that organises switching systems into logical collections, called peer groups. PNNI defines the creation and distribution of a topology database that describes the elements of the routing domain[1] as seen by an element. This database provides all the information required to compute a route from the given element to any address that is reachable in or through that routing domain. Path computation for each connection request comes from the information stored in these topology databases, including the connection traffic characteristics and requested Quality-of-Service. Connection request may demand complex combinations of Quality-of-Services elements. During connection set-up, each switching system along the chosen path performs Call Admission Control, which ensures the connection can be supported without jeopardising the Quality-of-Services guarantees to other existing connections. PNNI support a mechanism called crankback for partially releasing a connection set-up that has encountered a failure. The connection request is then going to be redirected along a different path.
An extension of PNNI is Integrated-PNNI (I-PNNI). I-PNNI provides a routing protocol for simultaneous support of IP packets and ATM SVCs in an IP (or Multi-Protocol) over ATM environment, based on PNNI. I-PNNI is a single routing protocol for both ATM and existing network layer protocols, such as IP. I-PNNI extends the capabilities of PNNI routing to the IP protocol, including Quality-of-Services routing and the ability to scale networks. It allows Quality-of-Services sensitive routes to be computed based on the complete end-to-end network topology. The work to specify this I-PNNI approach will be conducted by the PNNI Working Group of the ATM Forum.

3.7 Next Hop Resolution Protocol

The MAC-to-ATM address resolution provided by LANE involves two levels of resolution prior to data transfer. At first an internetworking layer address must be resolved to a MAC address, and then the MAC address is mapped to an ATM address. IP over ATM uses an enhanced resolution procedure that resolves internetworking addresses directly to ATM addresses. The disadvantage of standard resolution protocols is that they only takes place within a Logical IP subnetworks (LIS) or within a emulated LAN and not span the boundaries of a LIS or emulated LAN. The IETF has proposed a architectural model where the address resolution is extended beyond LIS boundaries, but within the ATM network boundary. The Next Hop Resolution Protocol (NHRP) provides this extended address resolution. The immediate advantage is reduction of the number of router hops to be taken by packets traversing from source to destination. The NHRP consists of two different components :

- ***Next Hop Server (NHS)*** - An entity that has the NHRP protocol server site implemented. Each NHS serves a set of destination hosts. The NHSs co-operatively resolve the next hop within the ATM network. The NHS maintains a cache, which is a table of address mapping for the IP to ATM-address resolution,
- ***Next Hop Client (NHC)*** - The initiator of a NHRP request.

[1] The routing domain is the logical grouping of nodes that share common peer group identifiers

In the NHRP model a physical ATM network is divided into Logical subnetworks, that in turn may consist of several LIS. Within an ATM-network direct connections are allowed, but ATM-networks that are separately administrated are connected via routers, even if belonging to the same ATM-network. figure 6 illustrates the principle of operation of NHRP. It shows an ATM-cloud with several subnetworks to what the NHRP extends. Each subnetwork has a NHS. Each station (a host or router) is associated with one or more NHSs. Typically the default router for the station is its NHS.

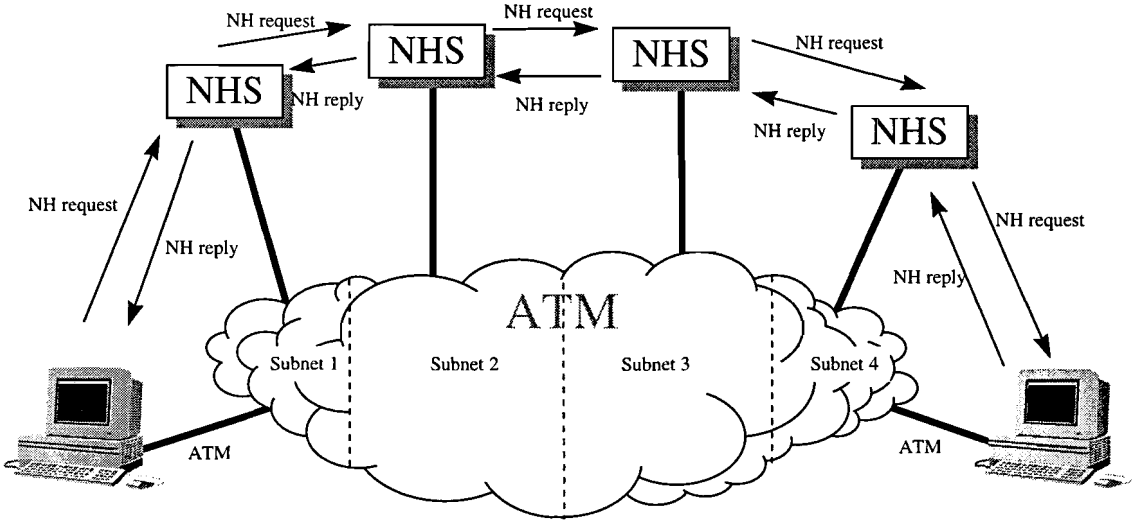


figure 6 : NHRP configuration

The NHRP defines the following steps when a source station (situated left in figure 6) has a packet for a destination station (right in figure 6) and the destination's ATM-address has to be resolved in order to set-up a connection.

1. The source station formulates an NH-Request packet containing its source IP- and ATM-address and the destination IP-address, and forwards the NH-Request to the NHS belonging to the subnetwork of the source station.
2. Upon receiving the NH-Request, the server checks to see if the destination IP-address is contained in its cache. There are two possibilities:
   A. there is no such entry. The NHS will forward the NH-Request to an other NHS closer to the destination which proceeds according to point 2;
   B. the NHS does have an entry for the destination, obtained in either of the following ways:
      I. it is the NHS of the destination ;
      II. the NHS has learned about the destinations addresses.
3. The NHS resolving the destination's ATM-address generates a NH-Reply packet. It can send the NH-Reply on behalf of the destination or based on its cached information to the source station. The packet may follow either of the following paths:
   A. the reverse path the NH-Request traversed;
   B. a direct connection to the source, if it exists.

The Next Hop Resolution Protocol allows a source station (host or router), wishing to communicate over a ATM network, to determine the IP and ATM address of the ATM next hop towards a destination station. NHRP can be used by a source station (host or router) connected to an ATM network to determine the IP and ATM network address of the "ATM next hop" towards a destination station. If the destination is connected to the ATM network, then the ATM next hop is the destination station itself. Otherwise, the ATM next hop is the egress router from the ATM network that is "nearest" to the destination station.

3.8 Multicast Address Resolution Server

Multicasting is the process whereby a source host or protocol entity sends a packet to multiple destinations simultaneously using a single operation. The more familiar cases of unicasting and broadcasting may be considered to be special cases of multicasting, where the packets are delivered to one destination, or to 'all' destinations. Most of the existing networklayer protocols, IP and IPX, assume that the underlying transport mechanism support multicast, for instance Ethernet or tokenring. ATM is being utilised as a link layer technology to support a variety of protocols. However, the ATM Forum's signalling specifications, UNI 3.1 and UNI 4.0, does not provide the multicast address abstraction. Unicast connections are supported by point to point, bi-directional VCCs. Multicasting is supported through point to multipoint unidirectional VCCs. The key limitation is that the sender must have prior knowledge of each intended recipient, and explicitly establish a VCC with itself as the root node and the recipients as the leaf nodes.

The Multicast Address Resolution Server (MARS) is an extended analogue of the ATMARP Server introduced in the IP over ATM environment, as mentioned in section 3.4. The MARS uses multicast Clusters. A multicast cluster is a collection of endpoints belonging to the same multicast address. The MARS may reside within any ATM endpoint that is directly addressable by the endpoints it is serving. Endpoints wishing to join a multicast cluster must be configured with the ATM address of the node on which the cluster's MARS resides. As mentioned before architecturally the MARS is an evolution of the ATMARP Server. Whilst the ATMARP Server keeps a table of *{IP,ATM}* address pairs for all IP endpoints in an LIS, the MARS keeps extended tables of *{layer 3 address, ATM.1, ATM.2, ... ATM.n}* mappings. It can either be configured with certain mappings, or dynamically 'learn' mappings. The format of the *{layer 3 address}* field is generally not interpreted by the MARS. A single ATM node may support multiple logical MARSs, each of them support a separate cluster. The restriction is that each MARS has a unique ATM-address. By definition a single instance of a MARS may not support more than one cluster. The MARS distributes group membership update information to cluster members over a point-to-multipoint VCC known as the ClusterControlVCC. Additionally, when Multicast Servers (MCSs) are being used, it also establishes a separate point-to-multipoint VCC to registered MCSs, known as the ServerControlVCC. All cluster members are leaf nodes of ClusterControlVCC. All registered multicast servers are leaf nodes of ServerControlVCC. The MARS does not take part in the actual multicasting of layer 3 data packets. It only resolves a multicast layer 3 address into several unicast ATM-addresses. See figure 7 for an example of a MARS.

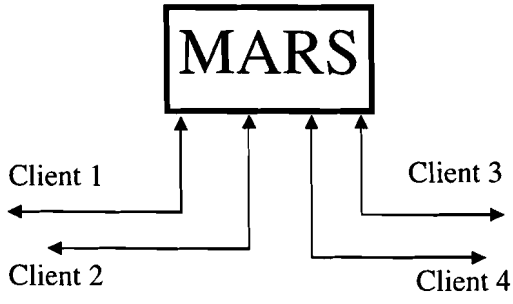


figure 7 : Example of a MARS

3.9 New Technologies

Several new technologies are currently in development to make network protocols faster then they are. IPv6 and Gigabit Ethernet are techniques to replace existing solutions. They only improve existing versions. So, IPv6 replaces IPv4 and Gigabit Ethernet replaces the underlying 10Mb/100Mb Ethernet connections. IP switching and Tag Switching are techniques that use the existing network protocols. IP switching and Tag-switching are using ATM as underlying transport mechanism to transport the network protocols more efficiently. An other new technical mechanism is the resource reservation protocol, which can deliver some kind of Quality of Service to the network layer.

3.9.1 IPv6

IPv6 or IP-next-generation (IPng) is a new version of IP which is designed to be an evolutionary step from IPv4. The reason why IPv6 is developed is the success of the predecessor, IPv4, and the explosive growth of Internet. As a result of this success several weaknesses are exposed. These weaknesses included the 32-bit address field, shortcomings in security and support of real-time traffic flows. The address field in IPv6 is 16 bytes long, in contrast to the 32-bit of IPv4. With this 16-bytes address field the problem in address space will be solved. Another improvement is the simplification of the header. It contains only 7 fields, versus 13 in IPv4. This change allows routers to process packets faster. IPv6 pays more attention to type of service by means of a flow label. This label indicates a properties and requirements belonging to a flow. Routers recognising this label can give packet a special treatment. With authentication and privacy, IPv6 takes care of security. To make IPv6 backwards compatible, address-space is reserved for IPv4 addresses and IPX addresses. Most vendors and industry analysts expect IPv6 upgrades by mid-1997 so it will be widely available.

3.9.2 Resource ReSerVation Protocol

IP provides best effort packet delivery that is sufficient for most of the conventional applications such as e-mail, WWW and file transfer. However, a new class of applications, like multimedia, is emerging and requires guaranteed resources from the network in order to function properly. The Resource reSerVation Protocol (RSVP) is enhancing IP based networks to support end-to-end quality of service, but it is however not related to ATM. The primary goal RSVP has set is : "efficient Internet support for applications that require service guarantees".

Protocol overview

Resource ReSerVation Protocol, [RSVP1],[RSVP2], has been proposed to be the protocol that allows applications to reserve network resources in an IP network such as the Internet. RSVP operates on top of IP (either IPv4 or IPv6) and it relies on standard Internet routing. It is used in both hosts and routers to reserve resources for a simplex (uni-directional) data stream, called a *flow*. A flow is a sequence of datagrams identified either by the IP destination address (either multicast or unicast address), or by the IP protocol ID and optionally by a destination port. The requested QoS for the flow is described by a *flowspec*

together with a *filter spec*. These two form a *flow descriptor* that is carried in the resource reservation message.

Filterspec.	Flowspec.	
(Specification of packet belonging to a flow)	TSpec. *(Data flow description)*	Rspec. *(QoS specification)*

Table A: RSVP Flow descriptor

RSVP is designed for both unicast and multicast communication in a heterogeneous network, where receivers may have different characteristics. These requirements lead to a solution, where the *receiver* is responsible for initiating the resource reservation.

The message flow for establishing of a network reservation for a multicast communication is shown in **figure 8**.
The sender S1 sends a Path message to a multicast group announcing the characteristics of the flow it is going to send. The Path message contains a Tspec, describing the maximum traffic characteristics of its data flow, and a Filter Spec, describing the packet format of the flow. When the receivers, R1 and R2, want to make a resource reservation, they will send a Resv message upstream following exactly the inverse path of the Path message. The Resv

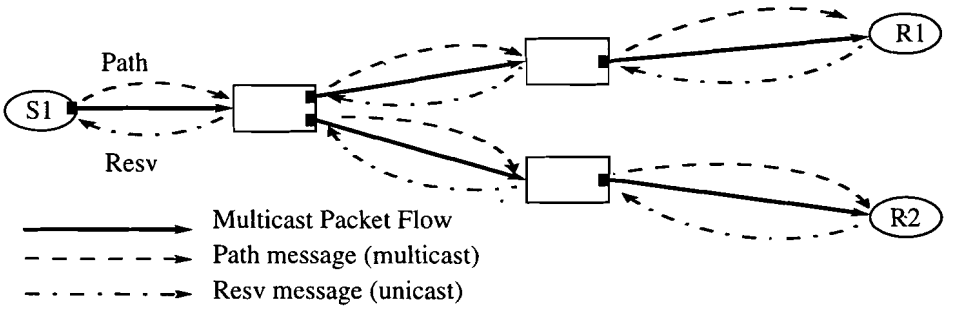


figure 8: RSVP message exchange in the multicast tree

message creates reservation state in each RSVP capable router along the path from the receiver to the sender. In a multicast situation, as the one shown in **figure 8**, there are nodes that will receive two or more Resv messages from different branches of a multipoint tree. These nodes merge the received reservations and forward only one merged reservation request upstream, containing the most demanding (maximum) flowspec.

3.9.3 Gigabit Ethernet

Gigabit Ethernet will provide 1-Gbps bandwidth for networks with the simplicity of Ethernet. This is because it employs the same CSMA/CD protocol, same frame format and same frame size as Ethernet. Gigabit Ethernet will be a backbone interconnect technology to be used between switches or high-performance servers. Gigabit Ethernet is being developed by the Gigabit Ethernet Task Force. This task force has as goal to complete the standard by 1998.

3.9.4 IP switching

IP Switching is a protocol designed and developed by Ipsilon Networks, and integrates IP routing software with high speed ATM switching in a single platform. The goal of IP Switching is to make IP work over ATM. Ipsilon's approach is to put an IP stack on an ATM switch. An IP Switch is the combination of a ATM Switch and a legacy router. Packet forwarding in a IP switch is handled by ATM switching and routing is performed by a

traditional router. IP Switching uses flow classification to optimise the IP switch. A *flow* is a sequence of IP packets sent from a particular source to a particular destination sharing the same protocol type (such as UDP or TCP), type of service, and other characteristics, as determined by information in the packet header. The IP switch identifies longer duration flows, as these can be optimised by cut-through switching in the ATM hardware. The rest of the traffic continues to receive the default treatment, hop-by-hop store-and-forward routing. End-to-end QoS can in principle be achieved in a homogeneous IP Switching equipped network. However QoS is only expressed with a priority for a flow and not with the usual ATM parameters for QoS.

Flow Classification

The main task of the flow classification process is to select those flows that are going to be switched in the ATM switch, and that should be forwarded sequentially by the router. The decision to switch flows directly through the ATM switch is called cut-through routing. Long duration flows are well adapted for cut-through routing. Short duration flows should be handled directly by the forwarding engine of the router. Application information provides an approximate indication for flow duration. Multimedia traffic (voice, image, video-conferencing) is an example of long duration flows, whereas name server queries are typically of short duration. For the flows selected for short-cut routing, a VCC must be established across the ATM switch and the association of flow and VCI label has to be communicated to the upstream IP switch to enable this switch to use a short-cut route. The Ipsilon Flow Management Protocol (IFMP) is used for the exchange of this information. There are 3 kinds of flows defined by the document [RFC1954] from Ipsilon.

Ipsilon Flow Management Protocol (IFMP)

IFMP enables communications between multiple IP Switches or between hosts and IP Switches. It associates IP flows with ATM virtual channels and defines the format for flow-redirect messages and acknowledgements. IFMP is implemented in end stations, such as routers, LAN switches, or TCP/IP hosts equipped with an ATM-card to connect directly to IP Switch. On ATM links it uses a default VCC (VPI=0, VCI=15). The ATM VCI for a specific IP flow is selected by the receiving end of the link. All packets of flows that have not been switched are forwarded hop-by-hop between IP switch controllers using the default VCC.

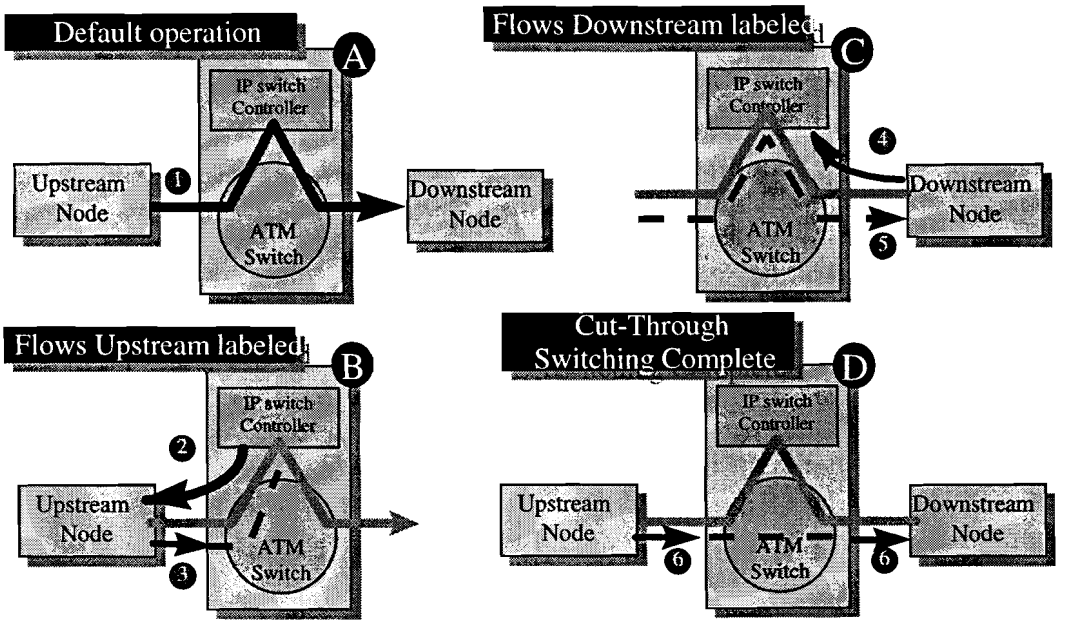


figure 9: IP switching Flow Management

At system start-up, each IP node sets up a virtual channel on each of its ATM physical links to be used as the default forwarding channel. An ATM input port inside the IP Switch receives incoming traffic from the upstream device on the default channel and sends it to the routing software of the IP Switch Controller (1) in **figure 9**. The IP Switch Controller forwards the packet in the normal manner over the default forwarding channel. It also performs flow classification, a decision-making process that enables IP Switches to optimise data traffic. Once a flow is identified, the switch controller requests the upstream node via IFMP to label that traffic using a new virtual channel (2). If the upstream node concurs, it selects a new virtual channel and the traffic starts to flow on this virtual channel (3). Independently, the downstream node can also request the IP Switch Controller to set up an outgoing virtual channel for the flow (4). When the flow is isolated to a particular input channel and a particular output channel (5), the IP Switch Controller instructs the switch to make the appropriate port mapping in hardware, bypassing the routing software and its associated processing overhead (6). This design allows IP Switches to forward packets at rates limited only by the aggregate throughput of the underlying switch engine. Further, because there is no need to reassemble ATM cells into IP packets at intermediate IP Switches, throughput remains optimised throughout the IP network.. More information about IFMP in [IFMP]

General Switch Management Protocol (GSMP)

The control protocol used between the IP switch controller and the ATM switch is the General Switch Management Protocol (GSMP). This allows IP switching to be used with ATM switches from different suppliers. GSMP is a simple master-slave, request-response protocol, and the switch issues a positive or a negative response, when the operation is complete. Unreliable transport is assumed between controller and switch for speed and simplicity. All GSMP messages are acknowledged, and the implementation handles its own retransmission. GSMP runs on the default VC (VPI 0, VCI 15). Five types of messages are involved : configuration, connection management, port management, statistics, and events. More information about GSMP in [GSMP].

3.9.5 Tag-switching

Tag Switching is a scheme for increasing the performance of IP. One way to does this is increasing the forwarding performance by routers. The need to improve forwarding performance while at the same time adding more routing functionality, allowing more flexible control over how traffic is routed, and providing the ability to build a hierarchy of routing knowledge.

Tag Switching consists of two components, namely forwarding and control. The forwarding component uses the tags carried by the packets and the tag forwarding information maintained by a tag switch to perform packet forwarding. The control component is responsible for maintaining correct tag forwarding information among a group of interconnected tag switches.

The control component is responsible for creating tag bindings, and then distributing the tag binding information among tag switches. Since the tag switch forwarding paradigm is based on label swapping, and since ATM forwarding is also based on label swapping, tag switch technology can easily be applied to ATM switches by implementing the control component of tag switching. The tag information needed for tag switching can be carried in the VCI field. To obtain the necessary control information, the switch should be able to participate, at minimum, as a peer in network layer routing protocols (e.g. OSFP, BGP).

When a packet with a tag is received by a tag switch, the switch uses the tag information as an index in its Tag Information Base. If the switch finds an valid entry it forwards the packet to the corresponding outgoing interface. The exact use of tag switching for QoS purposes depends a great deal on how QoS is deployed. If RSVP is used to request a certain QoS for a class of packets, then it would be necessary to allocate a tag corresponding to each RSVP session. For more information about Tag Switching see [Tag-switching].

4. Multi-Protocol Over ATM

Multi-Protocol over ATM, abbreviated Mpoa, is a technique to support network protocols over ATM. Mpoa makes it possible to transport these network protocols efficient, hereby bypassing the routers in the data path.

4.1 Introduction

The Multi-Protocol over ATM (MPOA) subwork-group of the ATM Forum has got the assignment to develop an standard approach to transport layer three protocols (thus explaining the name multi-protocol) over a ATM network. The goal of MPOA is to take advantage of the possibilities of ATM and at the same time offer the connectivity of an complete-routed environment. MPOA enables companies to implement ATM networks and let their layer three protocol take full advantage of the Quality-of-Service possibilities of ATM. One of the ways to explain MPOA is to look at it as a evolutionary step which takes you further then the LAN emulation specifications. MPOA operates on layer two and three, while LAN emulation only operates on layer two. Another difference between MPOA and LAN emulation is that LAN emulation only uses layer 3 subnetworks, while MPOA makes direct ATM connectivity possible between hosts of different subnetworks. Routers are required by LAN emulation to interconnect these subnets, but MPOA uses the Next Hop Resolution Protocol (NHRP) to allows intermediate routers to be bypassed on the data path. NHRP provides an extended address resolution protocol between different subnets. This enables the establishment of ATM SVC's across subnet boundaries, allowing inter-subnet communication without requiring routers in the data path.

MPOA is divided into two versions to come to its goals. MPOA version 1,which comes in its finally stage in July 1997, looks after the following functions :

- ATM Attached Hosts
- Next Hop Resolution Protocol (NHRP)
- Server Cache Synchronisation Protocol (SCSP)
- cache imposition
- trigger Protocol

The functions that remains for future MPOA version 2 are :

- MARS
- Multicast
- Quality of Service

Although the first version of MPOA may not be a solution for public networks, due to the absence of multicast and QoS, it is important to know what it does offer.

The following sections explain what MPOA is, give a description of MPOA, and the specifications of MPOA. The explanation of what MPOA is tells us where to position MPOA. The description of MPOA gives an outline of the MPOA protocol. It deals with the different components, the virtual router concept, configuration, discovery, target resolution, and in short data communication. The specifications of MPOA treat the component behaviour in more detail and specifies other protocols used by MPOA, like the keep-alive protocol and cache management protocols. Finally, data-communication is explained in more detail.

4.2 What is MPOA

MPOA allows network-layer protocols to communicatie between subnets over ATM without requiring routers in the data path. Future version of MPOA also let clients take full advantage of ATM's end-to-end QoS. MPOA also provides a framework for effectively synthesising bridging and routing with ATM in an environment of diverse protocols, network technologies, and IEEE 802.1 Virtual LANs. This framework is intended to provide a unified paradigm for overlaying layer 3 protocol on ATM. It provides direct connectivity between ATM attached devices in order to reduce latency and layer 3 processing load. MPOA is capable of using both routing and bridging information to locate the edge device closest to the addressed end station. This framework is composed of a number of protocols being developed in both the ATM Forum and IETF. MPOA integrates the NHRP work of the IETF and the LAN emulation work of the ATM Forum.

Additionally, the MPOA group developed solutions to the areas of this framework not already addressed by other bodies. Most notable among this category of solutions is the separation of route calculation from layer 3 forwarding, a technique which has come to be known as a virtual router, and is discussed in section 4.3.2.
The fundamental purpose of the MPOA Service is to provide end-to-end networking layer connectivity across an ATM fabric, including the case where some networking layer hosts are attached directly to the ATM fabric, some are attached to legacy subnetwork technologies, and some are using the ATM-Forum LAN emulation. Another aim of MPOA is to provide a good migration path from legacy LAN to a complete ATM network. For the setup of a connection, MPOA uses UNI-3.0, UNI-3.1 signalling and where possible UNI-4.0 signalling. Standard routing protocols of the networking protocol are supported by MPOA.

4.3 MPOA Description

MPOA consist of several logical components. Between those components there are different information flows. This section deals with the different kind of components and with the flows between them. The logical components consist of MPOA Servers (MPSs) and MPOA Clients (MPCs). MPOA performs the following operations, all discussed global in following part, the virtual router concept, configuration, discovery, target resolution, and in short data communication.

4.3.1 MPOA components

The ***MPOA Server (MPS)*** is a logical component of a router which provides network layer information to an MPOA System and resides within a router. It includes a full Next Hop Server (NHS) as defined in section 3.7 with some additional extensions.

- ***a router*** - a device that allows communication across subnetwork boundaries using a networking layer protocol. A router maintains tables for networking layer packet forwarding and may participate in one or more networking layer routing protocols for this purpose. A router forwards packets between subnetworks according to these tables. A router may contain one or more LAN interfaces, one or more LAN emulation Clients, and one or more MPOA Servers.
- ***a MPOA system*** - the set of inter-communicating MPOA-Clients and MPOA-Servers.
- ***a NHS (Next Hop Server)*** - Interacts with route protocols to propagate address - resolution queries.

The ***MPOA Client (MPC)*** is a logical component of an Edge-device or a MPOA host which provides network layer connections to other MPOA components.

- ***an Edge-device*** - a physical device capable of bridging packets between on or more LAN interfaces using one of the LAN emulation Clients (LECs). An Edge-device also contains one or more MPOA Clients allowing it to forward packets across subnetwork boundaries using a network layer protocol.
- ***a MPOA host*** - a host containing one or more LAN emulation Clients allowing it to communicate using LAN emulation. A MPOA host also contains one or more MPOA Clients allowing it to forward packets across subnetwork boundaries using a network layer protocol.

These functions are related to the forwarding of network layer packets and not to the operation of network layer routing protocols. The primary function of the MPC is to source and sink network shortcuts between subnets. To provide this function, the MPC performs network layer forwarding, but does not run network layer routing protocols. In its role, a MPC detects flows of packets that are being forwarded, using LAN emulation as default path, to a router that contains a MPS. When it recognises a flow that could benefit from a shortcut that bypasses the routed path, it uses a NHRP-based query-response protocol to request a shortcut to the destination. If a shortcut is available, the MPC caches the information, sets up a shortcut, and forwards frames for the destination over the shortcut. In its ingress role the MPC receives network data frames from other MPCs to be forwarded to its local interfaces/users. For frames received over a shortcut, the MPC adds the appropriate Data Link Layer (DLL) encapsulation and sends it to appropriate LEC. The DLL encapsulation information is provided to the MPC by a MPS and stored its cache.

4.3.2 Virtual Routers

A key aspect of the MPOA model is its need for virtual routers, a set of MPOA devices operating over an ATM fabric that collectively provide the functionality of a multi-protocol router. Since the edge devices accept data from an attached subnet, they are analogous to router interface cards. The ATM switching fabric can be seen as the backplane of the router, linking edge devices. The route server is analogous to the control processor.

The virtual routing approach makes it possible to deliver routing functions more efficiently and cost-effectively than today's routers can, resulting in edge devices that don't have to be as intelligent as a full-blown router. It also lead to more efficient scaling because adding forwarding capacity simply means adding switches, and adding additional routing capabilities means adding software to the route server. Management is easier too. The whole virtual router architecture, comprising multiple switches and route servers, can be managed as a single router. Finally, virtual routing allows for the creation of virtual subnets, since in a MPOA domain a hosts can physically be located anywhere in the network.

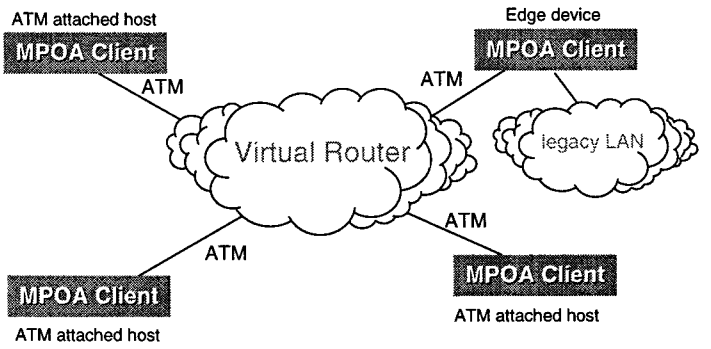


figure 1 : MPOA as virtual router

The separation of routing and forwarding will provide three key benefits :

1. it allows efficient inter-subnet communication,
2. it increases manageability by decreasing the number of devices that must be configured to perform network layer route calculation,
3. it increases scalebility by reducing the number of devices participating in network layer route calculation,
4. it reduces the complexity of edge devices by eliminating the need to perform network layer route calculation

4.3.3 Configuration

Before any communication, configuration has to take place. Configuration ensures that all components have the appropriate set of administrative information. All MPOA components require configuration information. By default, MPOA components retrieve their configuration parameters from the LAN emulation LECS. The MPOA components may be also obtain their configuration by other means. This information may be then provided by out-of-band mechanism, or by direct manipulation of the appropriate MIB allowing fine-tuning of the connections.

4.3.4 Discovery

Discovery is the process where the MPOA components learn of each others' existence and function. To reduce operational complexity, MPOA components may automatically discover each other using extensions to the LANE LE_ARP protocol that carry the MPOA device type (MPC or MPS) and the ATM address. This information is discovered dynamically and used as needed. This information may change and must be periodically verified.

4.3.5 Target resolution

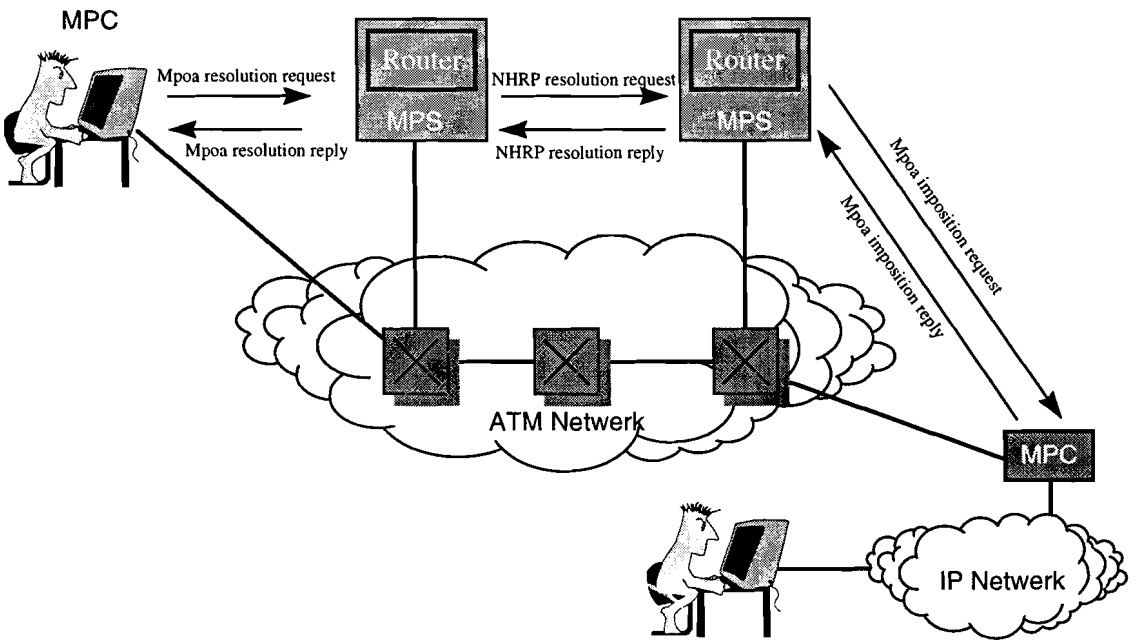


figure 2 : Target Resolution within a MPOA domain

MPOA Target resolution uses an extended NHRP Resolution Protocol to allow MPCs to determine the ATM address for end points of a shortcut. In the following part describes the protocol in several pieces. This is shown in figure 2, where the protocol is dived into six parts. Three parts from initiator side to destination and three parts the other way around.
The protocol starts with a *MPOA resolution request*. This *MPOA resolution request* is sent from a ingress MPC to request the ATM-address from the egress MPC, corresponding to a network layer destination address. Upon receipt of a *MPOA resolution request*, a ingress MPS must send a *NHRP Resolution Request* towards the egress MPS. Between the ingress- and egress MPS several MPSs may be located and its also possible that the ingress MPS and the egress MPS are physically the same. Upon receiving a NHRP resolution request, the egress MPS sends a *MPOA Cache imposition* request to the egress MPC. The path of a NHRP resolution pakket is out of the scope of MPOA.
The *MPOA Cache imposition request* is used to ask the egress MPC if it has the resources to except a shortcut by means of creating a egress cache entry. If the egress MPC has the corresponding resources to build a shortcut, a positive *MPOA cache imposition reply* is sent to the egress MPS. Via a *NHRP Resolution reply*, corresponding egress MPC information, including the destination ATM-address, is send back to the ingress MPS. Subsequently a *MPOA Resolution reply* is sent to the ingress MPC, with the corresponding egress MPC's information. The ingress MPCis now in possession of the ATM-address of the nearst location. serving the destination network address.

4.3.6 Example of a packets lifetime.

The most important part in MPOA is efficient unicast data communication. Unicast flows through the MPOA system has two primary modes of operation, the default flow and the shortcut flow. The default flow follows the routed hop-to-hop path over the ATM network. Per default, the MPOA edge device acts as a layer two bridge. Shortcuts are established by using the MPOA target resolution and cache management. When a MPOA Client has a

network protocol packet to send for which it has a shortcut, the MPOA edge device acts as a network layer forwarder and sends the packet over the shortcut.
There is a difference between inter- and intra-subnet communication. A intra-subnet flow is intended for a host on the same subnet. These packets don't benefit from a shortcut, because the data direct connection established through the defaut path follows the same path as a hosts on an subnet. Because intersubnet flows are always routed via a (default) router, there is a benefit using a shortcut.

As an example of data communication using MPOA, imagine a MPOA system that consist of MPSs en MPCs like Figure 3. Take into account that every MPOA component must be connected to the same ATM network, and that the drawn ATM subnets form a part of the complete ATM network. The MPSs are logically co-located with routers. All the components have been configured as defined in section 4.3.3, with information needed for correct system operation. A host (MPC 1) wants to communicate with a host (MPC 4) on a different subnet. The data-packet enters the MPOA system at a ingress MPC (1). By default, the data-packet is bridged to the default Router. If the data-packet follows the *default path*, it goes from router to router like ordinary network layer traffic via the ATM network. If no flow has been detected previously, each data-packet being sent to an MPS uses this *default path*. When a threshold, given as a number of packets for a single network layer address in a fixed period of time, is exceeded the MPC is obliged to use the Target Resolution process to obtain the ATM-address of the egress MPC (4). This ATM-address is used for establishing a shortcut to the egress MPC (4). If this packet is part of a flow for which already a shortcut is established, the MPC (1) strips the DLL encapsulation header from the packet and sends it via the Shortcut. When arriving via the Shortcut at the egress MPC (4), a packet is examined and either a matching egress cache entry is found or the packet is dropped. If a match is found, the packet is encapsulated using the DLL-information in the cache, and it is forwarded to the LAN interface. This LAN interface may be a bridge port, a internal host stack, etc. Appendix C provides some example scenarios for the data and control flows in a MPOA environment with one MPS. The presents of one MPS means that no translation between MPOA en NHRP took place. This can be caused by a ingress MPS which is at the same time the egress MPS.

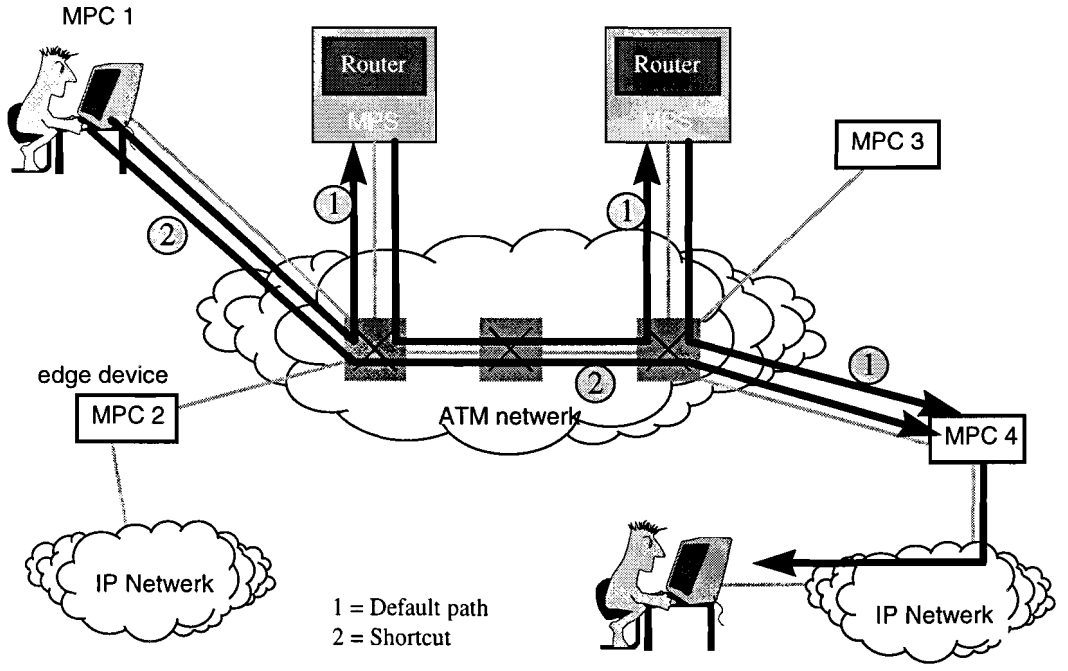


Figure 3 : Example MPOA data communication

4.4 MPOA Specification

This section specifies the behaviour of MPOA components in detail. First the use of LAN emulation in the default path will be explained and after that the MPOA Client behaviour is discussed. This includes inbound flow, outbound flow, flows within the same MPOA edge device, ingress- and egress cache management. Next, MPOA Server behaviour is discussed. This includes MPOA resolution and NHRP Resolution, and MPOA cache Imposition.
There are several protocols used by MPOA to provide robustness in the MPOA specifications. Those protocols include a MPOA keep-alive protocol that is used to check of a MPOA component still exists. Further there is a Cache management protocol that is concerned with the management of MPOA Client cache information by MPOA Servers. Finally there is a Data Plane Purge Protocol that is required to tell the ingress MPOA Client that a ingress cache entrie is no longer valid.

4.4.1 Default Path

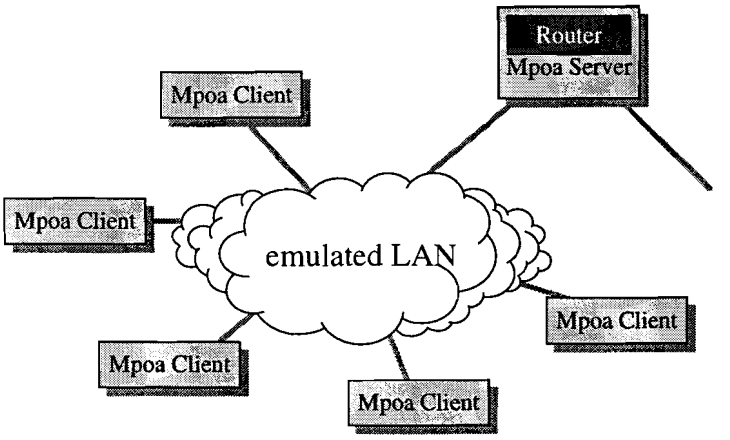


figure 4 : LAN emulation in MPOs default path

In order to support classical hosts that can be reached through edge devices, a MPOA system will need to make a particular virtual subnet look, to the classical hosts, like a single broadcast domain. That is, all of the edge device LAN, or WAN ports within a single virtual subnet would need to be bridged together. In order to do this, the MPOA protocol must interface with layer 2 subnet protocols, which provides this bridging function. The requirements of this protocol correspond closely with those of LAN emulation. This is why MPOA has introduced LAN emulation between MPOA Clients and MPOA Servers. It will also lead to a more natural evolution path from LAN emulation to MPOA. Each MPOA component will be equipped with a LAN emulation Client, so it can use the default path between MPOA clients and MPOA servers. Data-communciation as discussed in section 4.3.6 which follows the default path uses LAN emulation for the above mentioned reason. The whole situation is clearified in figure 4, this figure must be seen as part of Figure 3. The LAN emulation fundamentals are explained in section 3.5 or in [LANE].

4.4.2 Detailed MPOA Client Behaviour

The MPOA Client (MPC) lies between the LAN emulation Client and its Higher layers. Each LEC for which MPOA enabled, is associated with exactly one MPC. Each MPC serves a set of one or more LECs. The MPC presents the same service interface to its higher layers as its associated LECs present to them.
The MPCs maintain logical caches of information. These caches include mapping tables from target information to forwarding descriptions. There are two caches, used for incomming and for outgoing flows. These cache entries are aged by the MPCs according to the holding time specified by the MPS that provided it. The provider of a cache entry may retract it at any time, e.g. when routing changes or a station moves. A MPC analyses packets from the MPC

Service Interface for flow classification, collect statistics[1], and redirect packets to shortcuts. Non-direct packets are passed on to the LEC service interface corresponding to the MPC. Packets received from a LEC service interface are passed transparently up to the corresponding MPC service interface. The difference between a LAN emulation-capable bridge and a MPOA edge device is in the presence of a MPC. A MPC only deals with packets sent by the higher layers of the MPC service interface destined for a LEC, called the inbound flow. And the MPC deals with packets received on a shortcut service interface and relayed to the higher layers as if they came from a LEC, called the outbound flow. Figure 5 shows the difference between a LANE bridge and a MPOA enabled bridge.

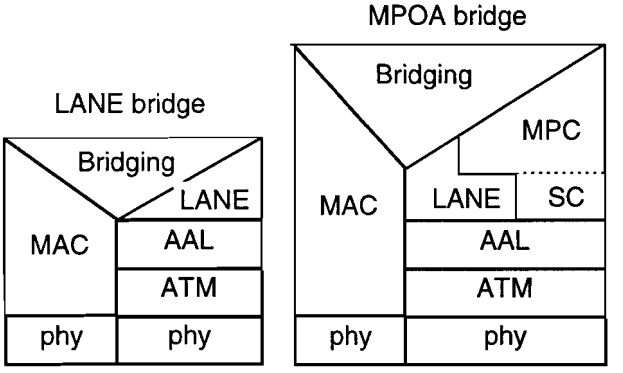


Figure 5 : Difference LANE bridge and MPOA bridge

Configuration

Configuration is used to ensure that all components have the right information to start up. Because of LAN emulation in the default path between MPCs and a MPS, the LAN emulation Configuration Server (LECS) can be used to distribute the configuration information. The MPC may send a configure request to the LECS to request MPC-specific configuration information. The configure request must contain MPOA device identification TLV's[2] identifying the LEC as a MPC. As a LECS accepts the MPOA device Identification TLV, it echoes it back in the response. Additional information override or initialise the corresponding values of MPS- or MPC-parameters. The ELAN-NAME/ELAN-NAME-SIZE fields in the Configuration Response should identify the ELAN for which the information applies. The following parameters apply to each MPS : Keep-alive time, and VCC Time-out period. Shortcut-Setup Frame Count, Shortcut-Setup Frame Time, VCC Time-out Period, and Flow detection Protocol are parameters applied to each MPC.

Discovery

It is necessary for the MPOA client to know the ATM-address of the MPOA Server so that a MPOA resolution request may be sent. Also, the MPS must know, if it issues a NHRP request, the ATM-address of a MPC so that a cache imposition may be sent. For the resolution protocol see section 4.3.4. To recognise each other, LAN emulation provides a convenient vehicle, by means of control messages like the *LE_ARP*.

Inbound flow

Inbound flow is data that enters the MPOA system and is send by higher layers[3] towards a LEC. All inbound packets are examined to see whether the have the destination MAC-address of a MPS. Together with the destination MAC-addresses the associated ATM-addresses is looked up in the LE_ARP cache, which is known through the above described discovery mechanism. If detection is enabled for the network protocol in the packet, the MPC examines the network layer destination address of the packet, and looks it in the ingress

[1] The MPC counts the passed packets

[2] Type/Length/Value : It is possible to add extra information to a LANE,MPOA and NHRP packets. To identify these extra bytes, you must indicate them by giving the type, the length and value of the extra information.

[3] As indicated in Figure 5 the bridging function also belongs to the higher layers.

cache for the *{MPS ATM-address, network layer address}* tuple. Flow detection and request, therefor, are on a *{MPS ATM-address, destination network address}* tuple basis. This lookup process can be modelled as a two stages process. First, a {LEC, MAC-address} to a MPS ATM-address lookup via the LAN emulation LE_ARP or corresponding LE_ARP caches. Secondly, a *{MPS ATM-address, destination network layer address}* to cache entry lookup. The contents of the ingress cache is shown in table 1.

Keys		Contents		
MPS destination ATM address	network layer Destination address	Destination ATM-address or VCC	Encapsulation information	Other information (holding time, count)

table 1 : Ingress Cache

In the absence of a cache entry, either because no cache table exists for the destination MAC-address or because a cache entry has not yet been established for this specific MPOA target, the data frame is forwarded to the appropriate LAN emulation Client. The absence of a MPS MAC-address indicates intranetwerk traffic. However the presence of a MPS MAC-address indactes internetwerk traffic which can benefit from a shortcut. If a MPS MAC-address is found, there are four possibilities, i.e.:

1. the *{ MPS ATM-address, network layer Address}* tuple is not found in the ingress cache.

A new cache entry is created in the ingress cache. In this entry the destination ATM-address/VCC field is not specified and the countfield is set to 1 to count the frame. Further the packet is send to the LEC interface for output on the emulated LAN. The entry is completely updated if a *MPOA Resolution Reply* is received after the thresshold is exceeded and a *MPOA Resoluiton request* is send.

1. the *{MPS ATM-address, network layer Address}* tuple is found, but the destination ATM-address/VCC field in the contents field does not specify a operational VCC.

The packet is counted in the Count field and send on to the LEC interface for output on the emulated LAN.

1. the *{MPS ATM-address, network layer Address}* tuple is found and threshold[1] is exceeded.

This indicates that a shortcut must be created. The MPC is then responsible for target resolution and sends a *MPOA Resolution Request* to the MPS to which the packet's MAC destination address is associated.

1. the *{MPS ATM-address, network layer Address}* tuple is found and the ATM-Address/VCC field is indicating a operational shortcut.

The packet header is stripped off and the packet is encapsulated with the appropriate network layer encapsulation. Then the packet is send over the specified shortcut.

[1] a preconfigured number of packets within a preconfigured time period.

Outbound flow

Outbound flow is data leaving the MPOA system. There are two ways that outbound flows can enter the MPC, i.e. via a LEC service interface or via the shortcut service interface. Packets received from the LEC service interface are passed transparently up to the corresponding MPC service interface. The MPC does not examine these packets but sends them to the higher layers. For all packets received on a shortcut, the MPC searches its egress cache for a matching entry. Such a entry is created as result of the acceptance of a MPOA imposition request. If the MPC receives this request it must determine whether or not it has the resources necessary to maintain the cache entry and the corresponding shortcut. A cache hit is defined as a match on the two main keys, *{network layer address, source/destination ATM-address}*, or one optional key, *{tag}*. The contents of the egress cache are shown in Table 2. If the received frame has a tag, this tag may be used to optimise the cache lookup,

Keys			Contents		
network layer destination address	source/dest. ATM-address[1]	Tag (optional)	LEC	DLL header	Other information (holding time, count)

Table 2 : Egress Cache

providing the same result as achieved through the main keys. In certain cases the tag information may be used to disambiguate multiple egress DLL headers on a given shortcut and network layer destination address. If an entry, matching a packet received on a shortcut, is not found in the egress cache, the packet is discarded, the error is counted, and the egress cache initiates a Data Plane purge, as discussed in section 4.4.5. If there is a cache hit, but the MAC destination address is not in the LE_ARP response tables in the LEC indicated by the egress cache entry, the packet is also discarded. Now the MPC will initiate a egress purge as described in section 4.4.6.

In the other cases where a cache hit takes place, and the corresponding LEC is fully operational, the DLL header in the egress cache is attached to the network layer packet. Then the resultant frame is passed to the MPC service interface as if it arrived from the LEC service interface.

Cache entry management

The ingress and egress caches are completely separate. Creation, deletion, or alternation of a entry in one cache does not imply any consequences for the other cache.

ingress cache entry management

An ingress MPC creates a new ingress cache entry when it detect a packet flowing to an MPS for an networking layer destination for which it does not already have a cache entry. The MPC initiates the Target resolution operation when the flow exceeds the threshold. When a MPOA resolution reply is received, the cache entry is completed.
Ingress cache entries are aged using the source holding time from the latest MPOA resolution reply received relative to the associated network layer destination address. Cache entries may also be withdrawn by the MPS at any time.
When an ingress MPC receives a MPOA Purge request it must stop using the shortcut for packets destined to the specified network layer destination address. It may issue a new MPOA resolution request immediately, or it may wait some time to send the query again. To prevent cache entries from ageing out, and to prevent that connections are severed, ingress caches could issue new MPOA resolution requests to refresh the cache entry.

[2] The source/destination ATM-address is needed as a key because it is possible for packets for a given network destination to be forwarded to different next hops based on where they came from.

egress cache entry management

When a MPS determines that it must impose an egress cache entry on a MPC, the MPS sends a MPOA cache imposition request to the MPC. The MPC uses the originating ATM-address and network layer destination address of the cache imposition as a key to find and/or create a egress cache entry. If a suitable shortcut from the originating ATM-address already exists, then the egress cache entry is attached to that shortcut, and can be used. If the shortcut does not yet exists, then it must be bound to the appropriate egress cache entry once it is created via signalling. The ingress MPC establishes the shortcut. Egress cache entries are created with a holding time provided by the egress MPS. The entry must not be used beyond this holding time. The egress MPS may refresh the entry if a appropriate resolution request is received, and thereby extend the holding time. An egress MPC may find that it must discard packets received over a shortcut because the egress cache entry may no longer be valid. If a MPC detects an invalid egress cache entry it must inform the MPS that imposed the egress cache entry by means of a egress cache purge. The MPS will then issue a NHRP purge request to the originator of the NHRP resolution reply for the network layer destination address. The MPS will delete its own association between the network layer destination address and the MPC ATM-address, so that the receipt of a subsequent NHRP resolution request will result in the relevant MPS finding out again the correct MPC ATM-address to use.

LAN-to-LAN flows within the same MPOA edge device[1]

An MPOA edge device can support multiple LAN interfaces. As a result, it is quite possible that a MPC needs to forward a frame between two MPC service interface without traversing a LEC and the ATM network. When creating an ingress cache entry, the MPC may recognise that the destination MPC is, in fact, in the same MPOA edge device. The MPOA edge device should then link the ingress and egress cache entries and perform network layer forwarding from the inbound MPC service interface to the outbound MPC service interface without transferring packets through a shortcut. There need to be a turn-around. There are several possibilities to redirect the data, i.e. in the inbound MPC service interface to outbound MPC service interface, or by a shortcut simulator, using the ATM hardware card, or at a ATM-switch.

[1] This is the same situation as if a edge device has two LECs, both connected to different emulated LAN, and were the LECs want to forward to each other. That can be valid, because the MPOA edge device acts as a layer two bridge, see 4.3.1.

4.4.3 Detailed MPOA Server Behaviour

A MPS is a component of a router. The data and control path from the LEC from a emulated LAN through the router is unaltered by MPOA. The MPS does, however, interact with the router, its LECs, the NHS, and other MPOA components. The router engages in the operation of traditional routing protocols like RIP and OSPF. One or more router interfaces may be used by LAN emulation. The MPS must be aware of the router configuration and forwarding tables to the extent of knowing whether a network layer destination address should be forwarded to a LEC.
The MPS must maintain the status of all ingress (by means of MPOA resolution requests) and egress cache entries (by means of MPOA cache imposition requests) that it has given to its MPCs. The MPS will generate responses and record the fact of responses. If the information becomes invalidated, a notification will go to the source of the Resolution Request. A destination may become invalid either because the actual host moved or expired, or due to a routing change. When the MPS has advertised its ATM-address via LE_ARP (the discovery protocol), it may receive MPOA resolution request from a MPC. Additionally, the MPS, in its capacity as NHS, it may receive NHRP queries from other NHSs. If the routing information indicates that the next hop is a MPC, then the resolution request is passed towards the MPC as a cache imposition request. Otherwise, the request should be treated as a standard NHRP resolution request and forwarded or answered according the NHRP protocol specifications. The MPS must maintain the status of all ingress (MPOA resolution replies) and egress cache entries (MPOA impositions requests) that it has given to its MPCs. The MPS will generate the reply, and record the fact of the reply. If any information becomes invalid, a notification will go to the source of the resolution reply.

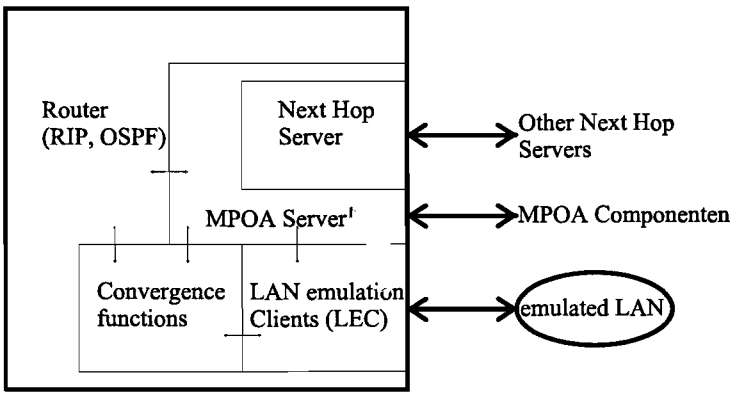


Figure 6 : Router and MPS

configuration

The MPS may send a configure request to the LECS to request MPC-specific configuration information. The configure request must contain MPOA device identification TLV's identifying the LEC as a MPS. The LECS on his turn may return only MPS TLV's in its response. MPS TLV's send by the LECS overrides the initial MPS parameters.

Interaction between MPOA Server and Next Hop Resolution Protocol

The role of a MPS in the NHRP can be descripted as a translator. In section 4.3.4 Target Resolution is discussed. The following section will discuss the same subject, but in more detail. First, the MPOA Resolution Requests to NHRP Resolution request translation is discussed. Next, the NHRP Resolution to MPOA cache imposition translation. Finally the MPOA egress cache replies to NHRP resolution Response translation and the NHRP Resolution Replies to MPOA resolution replies translating. The MPOA resolution request and replies are identical in format to the corresponding NHRP request and replies, except to fact that different packet types are used. Distinction is required because MPCs are assumed to be associated with edge devices, i.e. bridges to LANs. Specifically, since the MPC does not necessarily have a network layer address, the responding MPS, as NHS, may not be able to

deliver the reply to the ingress MPC, in absents of a Next Hop Client. So, MPOA requests are re-originated as NHRP requests. This re-origination ensures that the corresponding NHRP reply return to their point of origin, that is the ingress MPS.

1. *MPOA resolution requests to NHRP resolution request translation*
   When a MPS re-originates a MPOA resolution request to a NHRP resolution request, it creates a new request identifier for this NHRP packet. And forwards the NHRP packet to the next MPS with a NHS.

2. *NHRP resolution request to MPOA cache imposition request translation*
   When a MPS receives a NHRP resolution request from its NHS, it verifies of the router forwarding tables direct that network layer address to one of the LECs, known by the MPS. If so, the MPS communicates with the appropriate router convergence function, such as IP ARP, to determine the DLL header for frames send through the LEC to that destination. The MPS must subsequently check the LEC to see whether the LAN destination used to reach that network layer destination is served by a MPC. This information, along with the ATM-address of the MPC, is passed via LAN emulation LE_ARP control frame in the device type TLV, and is returned by the LEC to the inquiring MPS. Once this information is obtained, the MPS converts the NHRP request to a MPOA cache imposition request. In this request must be included all NHRP TLVs from the NHRP resolution request, emulated LAN identifiers, and the DLL header.

3. *MPOA cache imposition reply to NHRP resolution reply translation.*
   If the MPS fails to impose a egress cache entry at a MPC, then the egress MPS returns a Negative acknowledge or a NHRP reply with its own ATM-address.
   If a successful reply is received from the MPC, the MPS converts the successful MPOA cache imposition reply to a NHRP resolution reply. Furthermore, the egress MPS must maintain all valid unexpired MPOA cache imposition requests so that it may respond appropriately if a routing topology change occurs. If a imposition was successful, the egress MPS must maintain the mapping of network layer address to DLL header and ATM-address for the duration of the holding time.

4. *NHRP resolution reply to MPOA resolution reply translating*
   When the ingress MPS receives a NHRP resolution reply, the MPS converts this reply to a MPOA resolution reply and send this to the originator of the MPOA resolution request. The MPS construct the MPOA resolution reply with the original request ID, and source protocol address of the corresponding MPOA resolution request. Also, all other field are copied from the NHRP resolution reply.

4.4.4 Keep-alive Protocol

MPCs need to know that MPSs that have imposed egress cache entries are alive and able to maintain those egress cache entries. As such, the MPS is required to periodically transmit a MPOA keep-alive to all MPCs for which it has created and is maintaining egress cache entries. These must be sent once in a period specified by a MPS parameter.

4.4.5 Ingress cache maintenance protocol

The ingress cache entries are maintained by means of the following protocols: a MPOA trigger, a purge from the ingress MPS, and a data plane purge. A MPOA trigger is sent from a

ingress MPS to a ingress MPC, requesting the ingress MPC to issue MPOA resolution requests. A ingress purge is received by a ingress MPS to purge ingress caches for all relevant MPC, belonging to the purged destination address. A MPOA data plane purge is sent over a shortcut from a egress MPC to a ingress MPC to purge ingress cache entries.

MPOA trigger

A MPC must be able to detect inbound flows and establish shortcuts. In addition, a ingress MPS may detect inbound flows and request that the ingress MPC establish a shortcut. A trigger mechanism is used such that the rest of the resolution protocol remains consistent with the MPC-initiated mechanism. In the event that a ingress MPS determines the need for a shortcut, the ingress MPS may trigger the appropriate ingress MPC into initiating a NHRP Resolution Request. This is done using a MPOA trigger. The Ingress MPC, if it has the resources, responds by initiating a MPOA resolution Request for the target indicated by the MPOA trigger.

Purge from Ingress MPS

When a ingress MPS receives a NHRP purge request or a MPOA egress cache purge, it must send a NHRP purge request to all relevant MPCs for which it is maintaining ingress state for the purged destination address. The NHRP purge request are coming from upstream MPSs. The MPOA egress cache purge is received at the egress MPS from the egress MPC, and forwarded downstream to the ingress MPS as a NHRP purge request. If the received Purge contains a Client Information Element field, this may restrict the set of MPCs to which the purge must be forwarded.

Data plane purge protocol

A data plane purge is required to tell the upstream end of a VC that information it obtained via the MPOA variant of NHRP is no longer valid. A mechanism is provided to indicate that ALL cache entries associated with a particular shortcut should be purged. The different conditions under which a egress MPC is required to sent NHRP Purge request over shortcut are that a egress MPS has died or a egress cache Miss has occurred.

4.4.6 Egress cache maintenance protocol

A MPC must maintain the state for all the MPOA and NHRP resolution replies and successful MPOA cache imposition requests that it sources, for as long as the holding time is valid.. The holding time provided by the MPS is viewed as a contract in that the MPS guarantees, for the duration of the holding time, that whenever the information given to another party changes, it will send a notification to that party. This by an update or a purge. The recipient of the information is then free to use the information for the duration of the holding time. From the perspective of a egress MPS, the cache entries it maintains are those for which it has performed a successful MPOA cache imposition request and answered with a NHRP resolution request.

egress MPS purges and cache updates

When a MPS detects a change for a destination network layer address affecting one of its maintained cache entries, it must do one of the following two things. It either sends a NHRP

purge request to the set of affected sources of relevant resolution requests, or it must send a MPOA cache imposition request with a holding time of zero to the MPCs with the affected cache entries. There are several reasons why a change may occur, such as routing changes, bridging changes detected by routers, or a egress purge from a egress MPC.

egress MPC invalidation of imposed cache entries

An egress MPC may remove any imposed egress cache entry that has expired and should never use the information in such an entry until it is updated with a non-zero holding time. This updated information should by provided by the egress MPS that originally imposed it. Whenever a egress MPC receives packets one a shortcut, with a MAC destination address that is not know by one of its LECs, it must remove that cache entry and must send a MPOA egress cache purge request to the MPS that imposed that egress cache entry, see also section 4.4.5. When it receives a packet one a shortcut with a invalid egress cache entry, it must periodically send a MPOA data plane purge as described before in section 4.4.5.

MPC-initiated egress cache purge

The cache imposition protocol provides the capability for a egress MPC to issue a MPOA egress cache purge request, and for a MPS to issue a associated MPOA egress cache purge reply. When a MPS receives a MPC-initiated egress cache purge request it must first verify whether the change is due to a bridging change or a routing change. If the routing information is still unchanged, the MPS should verify the network layer address to MAC layer address mappings and MAC to ATM-address mappings, with for example a IP ARP and LE_ARP. Then the egress MPS generates the corresponding NHRP purge request and MPOA cache imposition requests. When a egress MPS receives a associated NHRP reply it issues a MPOA egress cache purge reply to the relevant egress MPC. If a shortcut is between a ingress and egress MPC, the NHRP purge request is sent to the ingress MPS that re-originated the NHRP resolution request after receiving the original MPOA resolution request from the ingress MPC. Then the ingress MPS forwards the NHRP purge request to the ingress MPC.

4.5 Data transfer

The primary goal of MPOA is the efficient transfer of unicast data. Unicast data-flow through the MPOA system has two primary modes of operation, i.e. the default flow and the shortcut flow. The default flow follows the routed path over the ATM network and the MPOA edge device acts as a layer two bridge. Shortcuts are established by using the MPOA target resolution and cache management mechanisms.
Shortly explained, edge devices examine the destination address of packets received on legacy LAN segments and decide how to forward such packets. If the packet doesn't need to go outside the network address summarisation group or virtual subnet, the edge device is finished. It merely bridges the packet, using *LAN emulation* to resolve the MAC-address to ATM address and establish a virtual circuit to the destination. In the other case, where a packet must leave its subnetwork, the MPOA Client checks the network address to check whether the packets must be sent to routers. A edge device can determine this, because those packets contains the MAC-addresses of router. If the packet must be routed, the edge device examines the contents of the packet to determine the destination network-layer address and looks up the ATM address corresponding to the destination network-layer address. The edge device then establishes a direct path over ATM, called a shortcut, to the appropriate destination.

The edge device receives the ATM address from either the route server or its own memory cache. The route server knows, or can use various routing protocols to discover, the ATM address of any device in the network. However, the design goal here is to minimise the number of visits the edge device must make to the route server to retrieve this information. Regarding this, the edge device maintains its own address caches. Much of the MPOA effort is devoted to devising effective cache-management techniques, including ensuring cache coherency between MPOA Clients and MPOA Servers.
Nowhere during this process the packet needs to be forwarded to a standard router. Instead, packet switching is handled by the edge device, while the route server performs address and routing resolution. A goal for the near future will be that this architecture can eliminate the scaling and performance bottlenecks described earlier.

If the local route server can not determine the appropriate ATM address, it can propagate the query to other route servers. The destination ATM address that the route server provides is either the address of the recipient host (if the host is ATM-attached) or the address of the edge device to which the non-ATM host is attached. When a MPC sends a network layer protocol for which it has a shortcut, the MPOA edge device acts as a network level forwarder and sends the packet over the shortcut.

A packet enters the MPOA system at the ingress MPC. The decision process that takes place relative to each inbound packet at a MPC is outlined in section 4.4.2. By default, the packet is bridged (layer two) via LAN emulation to the default router (this router doesn't need MPOA server capabilities). If the packet follows the default path, it leaves the MPOA system via the MPC's internal LEC service interface. However, if the packet is part of a flow for which a shortcut has been established, the MPC strips the DLL encapsulation from the packet and sends it via the shortcut. The MPC may be required to prefix the packet with tagging information prior to sending it via the shortcut. This tag is provided to the MPC via Target Resolution process as described in section 0. If no flow has been detected previously, each packet sent to a MPS is tailed by a network layer destination address as when it is being sent via LAN emulation. When a threshold is exceeded, the MPC is required to send a MPOA resolution request to obtain the ATM address to be used for establishing a shortcut to a specific downstream element, most likely a egress MPC. The MPS detect the requirements of a shortcut by means of the crossing of the threshold by the flow. The MPS subsequently sends a MPOA Trigger. The MPC will be notified of the need for a shortcut and starts the Resolution protocol.

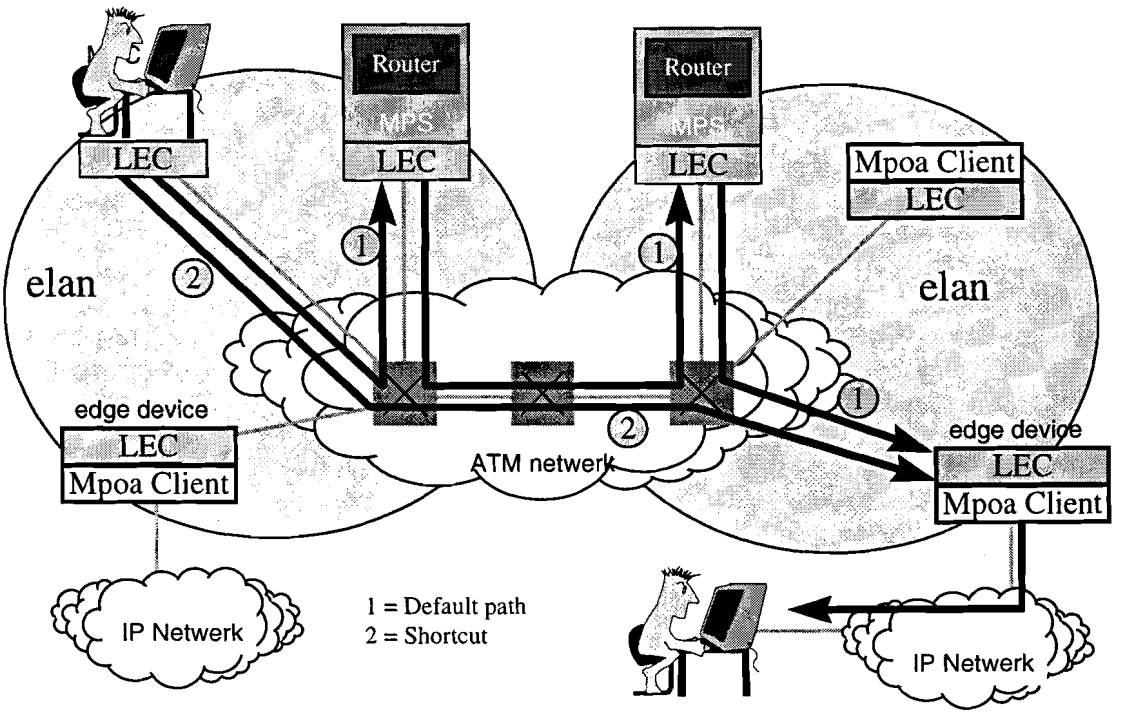


Figure 7 : Example MPOA data transfer

In legacy networks the traffic between two host follows a hop-by-hop path. To perform this the network layer uses a resolution protocol to find the hardware-address[1] of the router and afterwards it sends the packet. Summarised, the following packets are used by network layer traffic:

MPS destination ATM address	network layer destination address		Contents
(ingress cache table)			
network layer destination	**source/dest. ATM-address**		**Contents**
(egress cache table)			

1. ARP-packet to resolve the hardware address. (request)
2. ARP-packet with the resolved hardware address. (reply)
3. actual data packet send by the network layer to the below layer. Here the hardware address is added as extra information.

The above mentioned hop-by-hop sequence must be retained, with the only exception that the underlying medium doesn't use hardware addresses but ATM-addresses. The above packets are send to the MPC service interface by the network layer. In order to communicated to following steps take place :
When the ARP-packet is send to the MPC service interface, it is a inbound flow and the ingress cache is used to examine it. If this is the first time the host sends a packet, that is when there is no cache entry and no hit on the hardware-address in the cache of the MPS, the packet is send to the LAN emulation service interface. LAN emulation on his turn sends the ARP request through the BUS to every host connected on the emulated LAN. Because

[1] This hardware address is also known as the MAC-address.

the destination network address is on a different emulated LAN the router answers the request. The address resolution reply received by LAN emulation is transparently forwarded to the network layer. Subsequently the packet can be sent to the MPC's router. This process is repeated through the MPC service interface. Because this is still a inbound flow, the ingress cache is used to examine it. Now there is a MPS-MAC hit because the packet is send to a MPS and they will be resolved as result of the discovery process. The other key *{Network address}* in the ingress cache table doesn't give a hit, and a new entry in the ingress cache is created, with as contents an initial count of one and a invalid ATM/VCC field. Furthermore the packet is send through LEC service interface to the router following the default hop-by-hop path. Here the router forwards the packet over the other emulated LANs or LISs to the egress MPC's LEC.
From here, every packet designated for this network address is counted and send on the LEC service interface to the router following the default hop-by-hop path, until the threshold is reached. But if enough packets are send in a certain time, the threshold is exceeded. This indicates that a shortcut should be established. Via the target resolution, as described in section 4.3.5, the hardware address is resolved to the ATM-address and a shortcut can be established. The target resolution has as result that the ingress cache is updated and that a egress cache entry is created in the egress cache. From now on every packet will be sent through the shortcut. Appendix C provides some different example scenarios for data and control flows.

5. COMPARING MPOA WITH CLIP, LANE AND IP SWITCHING

This chapter discusses the comparison between techniques that deliver network layer protocols over ATM. These techniques are Clip, Lane, MPOA and IP Switching. The comparison is made on the following subjects : amount of network connections, geographical coverage, suitability for LAN or WAN, kind of generated traffic, kind of ATM connections, required services, QoS support, experience in the field and available products. Each section starts with a brief introduction and is divided into several subsections specific to each subject. More attention is paid to the coupling between the Resource reSerVation Protocol and ATMs QoS, in section 5.7. This chapter doesn't explain the working principle, as this is already done in section 3.9.2. After an introduction on the different flows used in RSVP and the translation of them into ATM VCCs, it gives a realistic view of the use of RSVP in combination with Clip, Lane, Mpoa, and IP Switching. This chapter ends with a survey on the pros and cons of Clip, Lane, Mpoa, and IP switching. A conclusion of the result of work discussed here is given in chapter six, about the interest of using Mpoa or the other protocols.

5.1 AMOUNT OF NETWORK CONNECTIONS

Clip, Lane, Mpoa, and IP Switching use ATM as underlying transmission medium. In ATM a virtual connection, called VC, is used to connection two clients or components. This section looks at the amount of VCs used by each technique to support the technique by looking after the number of VCs used for data transport. Each techniques uses some components that forms the core of the technique. For Clip the core function is the IP-ATM-ARP server, in Lane the Lane services : LES, BUS and the LECS. In Mpoa the main functions are performed by Mpoa Servers. In IP switching the core functions are the IP switches. There are several ways to connect these functions. This is shown in figure 1, where the different ways to connect these core components are explained.

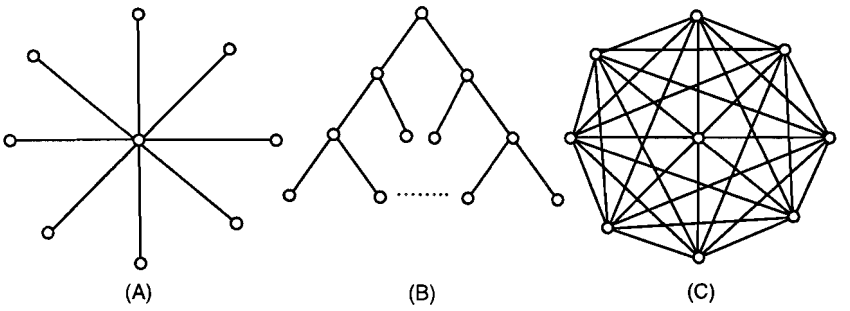


figure 1: Some possible topologies for point-to-point connections. (A) Star. (B) Tree. (C) Complete.

A. In a star topology there are n-1 connections. This gives 2(n-1) VCs in an environment with only uni-directional point-to-point VCs and n-1 VCs in an environment with only bi-directional point-to-point VCs.
B. A tree gives n VCs in an environment with only uni-directional point-to-point VCs or 2n VCs in an environment with only bi-directional point-to-point VCs.
C. In a complete network, where every node is connection with every other node and in an environment with only uni-directional point-to-point VCs, there are $n(n-1) = n^2-n$ VCs.
In a complete network, where every node is connection with every other node and in an environment with only bi-directional point-to-point VCs, there are $\frac{1}{2}(n^2-n)$ VCs.

In the following subsections the assumption is made, that bi-directional point-to-point VCs are supported and that uni-directional point-to-point VC are only used if necessary.

5.1.1 CLIP

In Clip the IP-ATM-ARP is the core function. Each client in Clip is connected to this service. This situation is like a star-network. In a network with n clients, **n** bi-directional point-to-point VCs are needed to connect the clients to the IP-ATM-ARP server.

5.1.2 LANE V1.0 AND LANE V2.0

As mentioned before Lane has a service that consist of three components, the LES, BUS, and LECS. Lane v1.0 needs **3n + 2** VCs to connect n LANE clients (LEC) to the Lane services. This is subdivided in 3n VC per Client, one to the LECS, one to the BUS and one the LES. Further, Lane has specified that a LES and BUS are also connected to each LANE Client with a point-to-multipoint VC. This adds the total needed number of VC for Lane v1 to 3n+2. Summarised :
3N : Three point-to-point VCs per LEC (n) to the LECS, LES, BUS.
2 : Two point-to-multipoint VCs, one from the LES and one from the BUS to each LEC.

Lane v2.0 needs $\mathbf{3n + 2s + \frac{1}{2}(s^2-s) + 2s}$ connections to connect n Lane Clients to the Lane Service. This is subdivided as follows :
3n : Three point-to-point VCs per LEC (=n) towards the LECS, LES, and BUS.
2s : Two point-to-multipoint VCs from each LES/BUS combination (=s) towards each LEC.
$\frac{1}{2}(s^2-s)$: Two point-to-point VCs with a complete structure between each LES/BUS combination.
2s : Two point-to-point VCs between each LES/BUS combination towards LECS.

5.1.3 MPOA

The Mpoa Service consists of Mpoa clients and Mpoa Servers. Mpoa consist of a default path and a shortcut path. The default path uses Lane v2.0 and requires $3n+2s+\frac{1}{2}(s^2-s)+2s$ VCs, as calculated in section 5.1.2.
Additional Mpoa itself needs $\underline{n+\frac{1}{2}(s^2-s)}$ VCs to work correctly.
This is subdivided as follows :
n : Each of the n MPOA Clients has a point-to-point VC to his MPS (there could be more MPS).
$\frac{1}{2}(n^2-s)$: A complete structure between the s MPS (MPOA Servers).

It is possible to set-up a separated VC to transport the MPOA triggers from a MPS to a MPC, this is not take into account in the above calculation.

5.1.4 IP-SWITCHING

IP Switching specifies a default connection between each IP switch and components. The components include IP Switch gateways, multihomed routers, and IP switch attached hosts. To connect **n** IP components to one IP switch there n VC are needed. Each IP switch itself is connected to another IP switch. If they are complete connected there are $\frac{1}{2}(s^2-s)$ VC needed. Totally $\underline{n+\frac{1}{2}(s^2-s)}$ VCs are needed.

5.1.5 CONCLUSION

Technique	Amount of network connections
Clip	n
Lane v1	3n + 2
Lane v2	$\{3n + 2s + \frac{1}{2}(s^2-s) + 2s\}$
Mpoa	$\{3n+2s+\frac{1}{2}(s^2-s)\} + \{2s + n+\frac{1}{2}(s^2-s)\}$
IP switching	$n+\frac{1}{2}(s^2-s)$

Clip uses the least amount of network connection. Mpoa uses the most. The reason is that Lane is used in Mpoa servers and Mpoa clients in the default path.

5.2 GEOGRAPHICAL COVERAGE

This subchapter deals with the scaleability in respect to the amount of routers and clients, the size of the network and the performance aspects.

5.2.1 AMOUNT OF ROUTERS

Clip and Lane are techniques that uses routers for communication between LISes and elans. When there are many routers in the data-path the end-to-end delay can have a major influence on the traffic performance. A question to answer first is "Why does one need routers?". As mentioned in section 3.4 and 3.5 a LIS, local IP subnetwork, and an elan, emulated LAN,

allows users to be in the same IP subnetwork, this means that all the clients in a lis or elan must have the same IP network address. It is not realistic to give all the existing users the same network address. For example this gives huge IP-ATM-ARP tables for Clip and enormous amount of broadcast traffic in Lane. Besides all this, large address sizes that are necessary to separate clients.

The reasons why someone use routers or why clients are grouped together in a LIS or ELAN are :

1) Current legacy LANs are already build into subnets.
2) to make communication possible with legacy LANs through routers.
3) to separate broadcast domains
4) to screen unknown traffic
5) to protection considerations
6) for management and control
7) to couple independent networks

In an environment with one or more routers Mpoa and IP switching have performance advantage, because the router can be avoided. Because Mpoa and IP switching in most cases set-up a connection that will follow the same path. A choice only based on "amount of routers" can not be made. One must also take other selection criteria into consideration, like the support of many clients, the size of the network, network performance, etc.

5.2.2 AMOUNT OF CLIENTS

In an intra-network with a lot of clients and no bridges, Clip is the solution above Lane. This solution has lesser overhead compared to Lane, and Clip supports a greater MTU size. Clip is a simpler protocol and is faster. A disadvantage is that Clip does not support multicast and broadcast traffic. In a subnetwork with a lot of clients who are spread over a great distance it is a lot more efficient to place more servers in the subnetwork. More servers in a subnetwork makes it possible to increase broadcast traffic and can also increase the distance between a client and server. Lane v2 does support this by allowing more LES/BUS pairs. The amount of clients supported in Mpoa is the same as with Lane, because Mpoa uses Lane in the default path. The amount of clients in IP switching depends on the support of IP switch gateways, because IP switching is concentrated at the routers bottleneck and not on the client side.

5.2.3 SIZES OF NETWORKS

As before mentioned in section 5.2.2, it is recommended to use Lane above Clip in a network with great distances between clients, because Lane support multiple LES/BUS pairs. With Clip it is also possible to use more Servers, but they are not tuned with each other like with Lane. In a network with a small amount of clients, for example four clients, its is recommendable to use Clip, because simplicity wins from complexity of Lane. A situation like this is an ATM backbone, which connect a small amount of routers together.

5.2.4 DELAYS AND NETWORK PERFORMANCE

Occurring delays and network performance can only be determined within a test environment. At KPN research several tests are already done. Network performances in a Clip environment with Permanent VC is done, and a performance test of Lane v1 is also

done. To come to the correct conclusion also several tests must be done, on Lane v2, Mpoa, and IP switching.

Before one makes a performance test, one must make a theoretical analysis. With this analysis the following parameters must be determined :

- **Network delay** - Minimum time it takes for a packet to traverse the network end-to-end, and consist of the following parts :
    - * Input delay - A router must completely buffer a packet before it can be sent on.
    - * Queuing delay - Delay caused by different inputs designated for the same output.
    - * Fixed router delay - the fixed store and forward delay wasted in a router.
    - * Propagation delay - 5μs/km
- **Receive buffer delay** - delay introduced by the play-out buffer which compensates the delay variations, also called jitter delay.
- **Fill delay** - consists of two parts
    - * Coding delay - The time it takes to code data.
    - * Packetisation delay - The time to fill a ATM-cell or IP-packet. (calculated by dividing the effective packet/cell size by the information rate. An example calculation in a situation with an information rate of 64 kbit/sec over ATM levers a delay of (48byte/cell * 8bit/byte) / 64kbit/sec = 6 msec.

After the theoretical determinations of the above parameters a filed test should be done. This is for future experiments.

5.3 SUITABLE FOR LAN OR WAN

When a comparison is made based on whether or not it is suitable for LAN or WAN, one must know what is meant with LAN or WAN. A LAN, local area network, is a private network within a single building, campus, or in general inside an area of a view kilometres. A LAN network distinguishes itself from other networks on three points, i.e. the size, topology, and transmission technology such as Ethernet or token ring/bus. A MAN, Metropolitan Area network, is bigger then a LAN. The major point of a MAN is, that it is a broadcast medium, where all the computers are connected, for example with DQDB, Distributed Queue Distributed Bus. A WAN, wide area network, spans a wide geographical area, which could be a country or even a continent. A WAN consists of hosts, end-stations which are connected by means of routers. A LAN always uses a symmetrical topology, Ethernet or tokenring. But in a WAN several topologies are possible, such as star, ring, tree, complete, intersecting rings, or an arbitrary connection.

This sections starts with the first impression of when to use Clip, Lane, Mpoa or IP switching. After that interconnection is discussed. Interconnection is used to connect networks together. Here several interconnection functions are treated such as Network service, Addressing, Routing, Address resolution, and the cost of interconnection.

5.3.1 FIRST IMPRESSION

The first impression on where to use each technology is shown in table 1.

Interprocesor distance	Position Computers	Network type	Suitable
<10 m	room	LAN	LANE/CLIP
<100 m	building	LAN	LANE/CLIP
<1 km	Campus	LAN	LANE, CLIP, ~~MPOA, IP switching~~
<10 km	city	MAN	~~LANEv1~~, LANEv2, CLIP, MPOA, IP switching
<100 km	Country	WAN	LANEv2, MPOA, IP switching
<1,000 km	Continent	WAN	MPOA, IP switching
>1,000 km	Planet	Internet	MPOA, IP switching

table 1 : Suitable for LAN, MAN, or WAN

5.3.2 INTERCONNECTION

Clients of different networks should be able to communicate with each other. To enable this, some kind of interconnection has to take place. Some of the technical aspects of interconnection are :

1) Specification of the interface for physical connection.
2) Support of transmission and signalling.
3) Switching capacity.
4) Support of advanced technologies, such as address portability.

In short, an interconnection point is a point where a translation takes place between two different types of networks. This point is called "point of interconnection" (POI) and normally takes place within a bridge, router, or gateway, because these are devices that take care of the translation.

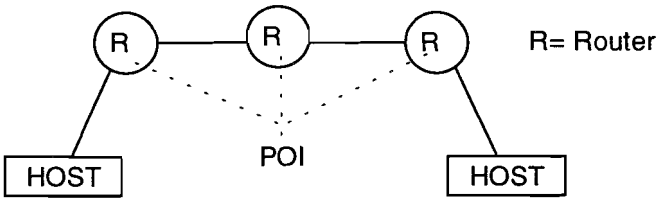


figure 2 : point of Interconnection

Interworking within ATM network, is related with the following subjects [Voogt]: Network services, ATM-addresses, routing information and address resolution, and many more subjects. These subject are called the interworking functions. These function are described below in detail according to the used technique.

Network services

The network service pays attention to the support of different networks and the support of different network protocols.

Clip is a layer three protocol, and as the name says it only supports IP networks. Other network protocols, i.e. IPX or Apple Talk are not supported. Nowadays two versions of the IP protocol exist, namely IPv4 and IPv6. There is no research done on the interconnection between two LISses with both version of the IP protocol. A router or gateway should take care of the translation between those LISses. Lane is a layer two protocol and has to connect layer two protocols. Lane supports layer two bridges which takes care of translations between legacy LANs and the ATM-network. For the support of different legacy LANs Lane must specify different elans. Each elan supports clients of the same LAN. For example an elan may only have Ethernet clients or token ring clients.

Mpoa works on two layers, layer two and layer three. Mpoa supports both bridges and routers. Bridges and routers both take care of translation between different subnets. Beside IP protocols Mpoa supports also IPX, DECnet, AppleTalk and other network protocols, but no interconnection is defined between them. A higher layer interconnection point like a gateway has to resolve the translation as point of interconnection. IP switching is a layer three protocol like Clip, that is capable of supporting one network protocol, in this case, the IP protocol.

Addressing

Addressing pays attention to the use of different type of ATM-addresses. Within Clip there are no problems with the translation of IP-addresses to ATM-addresses, as long as the routing protocols support it and the IP-ATM-ARP server can resolve them.
The Lane specification doesn't say anything about the support of different types of ATM-addresses. For both Lane and Clip the problem of supporting ATM-networks with different ATM addressing is not solved yet. A router should be used to connect them. Mpoa has no problem in supporting different ATM addressing, because it supports routers. On the point of addressing, IP switching is the best, because the ATM addressing and signalling is not used at all.

Routing

Routing is used to determine the route of a packet, by exchanging routing information. Clip, Lane, Mpoa, and IP switching look at routing information as data traffic, and treat it likewise. Lane is a layer two protocol and does not use network routing protocols, it only uses ATM signalling and routing to find a client on the ATM-network. Network routing is necessary where two elans need to be connected. Clip uses two kind of routing protocols, the traditional network routing protocols and the ATM signalling and routing. The difference in use is that the traditional routing protocols are used for connecting LISses whereas the ATM signalling and routing is used within a LIS. Thus routing and ATM signalling are used on a different hierarchical level. Mpoa is not like Clip on this point, it uses the traditional routing for the default path and ATM signalling and routing on the shortcut path. They are used next to each other. IP switching does not support ATM signalling and routing but only uses network routing.

Address resolution

Address resolution is used to resolve addresses. Address resolution in Clip is resolved by a special IP-ATM-ARP server. Address resolution in Lane is resolved by a LES and unknown traffic by an BUS. Lane supports the use of more then one LES/BUSses in one elan, to spread the ARP traffic. In Mpoa the address resolution in the default path is solved by the Lane. For the shortcut address resolution an apart protocol is defined, which is an extension of the NHRP protocol. IP switching doesn't use address resolution.

Costs of interconnection

The cost of interconnection depends on :

1) the amount of components used for interconnection.
2) the need of hardware/software upgrades.
3) the complexity.
4) the scaleability.

No concrete cost can be given on the moment, because Lane v2.0, Mpoa and IP switching are still a draft version.
More complexity may give an improvement of the scaleability, but expands the costs. Lesser complexity gives lesser scaleability but lower costs.

5.4 KIND OF GENERATED TRAFFIC

This subchapter discusses the kind of control traffic generated by the different protocols and gives an indication on how often they are generated. The packets generated by the ATM signalling for VC set-up is not included here. In every section discusses the kind of control traffic per situation.

5.4.1 CLIP

Clip does not specify any configuration traffic. Every Clip component has to be set-up by hand. Also the registration is done manually. For address resolution two packets are necessary. One request and one response. These two packets are only generated when a client wants to known the ATM address of a destination belonging with the IP address. For data traffic no special control packets are used and also for the maintenance of data connection no special packets are used.

5.4.2 LANE

Lane has specified two packets for configuration traffic. Every client has to send a configuration packet to the Configuration server at start-up. The configuration server on his turn answers the request. After the configuration the client has to register it self by the Lane Server. Two packets are used for the registration, a request and a response. At start-up also two LE_ARP packets are generated to find the BUS. For address resolution from MAC-address to ATM-address two packets are used, again a LE_ARP request and a LE_ARP response. For data traffic, one packet is used to indicate that the data channel is available. For maintenance of the data channel no packets are used.

5.4.3 MPOA

Mpoa uses Lane for configuration and registration of the LEC inside a MPC or MPS, thus uses the same amount of packets. This packets are extended with a so called TLV-field to indicated a client as a Mpoa Client or as Mpoa Server. For discovery of each component Mpoa uses the LE-ARP protocol. Mpoa consists of two data paths, a default path and a shortcut path. The default path follows the hop-by-hop path and uses Lane. For the address resolution to set-up a shortcut, Mpoa uses six different packets. Two packets between the ingress MPC and ingress MPS, two packets between the MPSses, and two packets between egress MPS and egress MPC. These packets are, in the same sequence, *Mpoa resolution requests/replies, NHRP resolution requests/replies,* and *Mpoa imposition requests/replies.* These packets are uses once per set-up of a shortcut. For data traffic on a shortcut no additional control traffic is generated. For maintenance the same packets are used as for the address resolution. These packets are generated according to the holding time of a cache entry, which has a default of 20 minutes. Maintenance can take place in five different ways, depending on the initiator. The packets used for maintenance are called purges and is discussed in section 4.4.5 and 4.4.6 where the Mpoa cache management is discussed.

5.4.4 IP SWITCHING

IP switching is a hop-by-hop protocol. This means that only adjacency IP switches are aware of each other existence. For synchronisation between two IP switches an handshake protocol is used. This requires three packets, a syn, a syn_ack, and an ack[1]. Because a complete IP switch consists of two parts : a router and a switch. Synchronisation between those two components also has to take place. This requires eight packets. One of the advances of IP switching is that no address resolution has to be done on layer three. For the set-up of a cut-through path two packets are used, i.e. a IFMP redirect message and a reply. For the maintenance of the cut-through path the same two packets are needed. The update takes place once every twenty minutes.

5.5 KIND OF ATM CONNECTIONS

5.5.1 CLIP

For signalling within Clip UNI is used via VCC= 0,5. Further every client needs a connection to the IP-ATM-ARP server. Further no additional VCCs are used. Between clients one or more uni-directional unicast VCC can be set-up.

5.5.2 LANE V1.0

For signalling within Lane UNI is used via VCC = 0,5. Other control VCCs are :

- Between client and Configuration server, one bi-directional unicast VCC.
- Between client and Server, one bi-directional unicast VCC and one uni-directional multicast VCC set-up from the server.
- Between client and BUS, one bi-directional unicast VCC and one uni-directional multicast VCC set-up from the BUS.
- Between clients, one or more uni-directional unicast VCC

5.5.3 LANE V2.0

Lane v2.0 has the same kind of ATM connections, the only difference is the existence of multiple LES/BUS pairs. These LES/BUS pairs are mutually connected by a bi-directional unicast VCC. Because Lane v2.0 specifies the use of a I(nteligent)-BUS or Special Multicast Server (SMS), additional VCCs are required. The I-BUS replaces the current BUS. The SMS requires a bi-directional unicast VCC from every client. And the SMS sets up one uni-directional multicast VCC per multicast address.

[1] Syn is an abbreviation of Synchronise
Ack is an abbreviation of Acknowledge

5.5.4 MPOA

For signalling within Mpoa UNI is used via VCC = 0,5. It is possible to use I-PNNI, but this is not specified. Further control VCCs used by Mpoa are the following :
Configuration information is provided by the Lane v2.0 connections to the LECS.
Discovery information is provided by the Lane v2.0 connection to the BUS.
Target resolution consist of two parts the Mpoa resolution and NHRP resolution :

- Mpoa resolution from MPC to MPS, a MPC initiated VCC is used in combination with LLC/SNAP encapsulation. The Mpoa resolution packets aren't sent by Lane because the uses of Lane is transparent for the MPS.
- NHRP resolution between MPSs. Because of the same reason as by Mpoa resolution, a separate VCC is set-up for the NHRP resolution and also LLC/SNAP encapsulation is used.

Keep-Alive packets are send from MPSses towards multiple MPCs. A separate VCC can be used or the same VCC as for target resolution can be used in combination with LLC/SNAP encapsulation.
Purges exist there in several types, depending on the initiator of the purge.

- Mpoa trigger from MPS towards a MPC, via an unicast VCC with LLC/SNAP encapsulation.
- Mpoa egress cache purge from MPC towards MPS, via an unicast VCC with LLC/SNAP encapsulation VCC. This purge is translated into a NHRP purge.
- NHRP purge between MPSs, via an unicast VCC with LLC/SNAP encapsulation VCC. This purge is translated into a Mpoa imposition request.
- Data plane purge from MPC to MPC. This purges uses the shortcut between two clients.

Routing Information is seen as normal data traffic and is send through Lane or Clip.
Data traffic. In Mpoa two kind of data connections can be used, the default path and the shortcut.
For the default path the following applies :

- The connections between ingress MPC and ingress MPS are provided by Lane v2.0. The connections between ingress MPS and egress MPS are provided by Lane v2.0 or Clip. The connections between egress MPS and egress MPC are provided by Lane v2.0.
- For the shortcut a bi-directional unicast is set-up.

5.5.5 IP-SWITCHING

IP switching has no reserved VCC beside the (0,0) for the unassigned cell and the (0,15) for the use of default encapsulation. Each packet send by IP switching initially uses the default encapsulation. If a flow is detected for a packet, a VCC is reserved for that flow. The VCC isn't set-up by the ATM's UNI but by the IFMP protocol.

5.6 REQUIRED SERVICES

This subchapter discusses the services that are required by Clip, Lane, Mpoa, and IP switching. Hereby dealing with the used kind of encapsulation and the support of Multicast traffic and which ATM VCs are used or reserved.

5.6.1 ENCAPSULATION

Encapsulation is the method of how traffic is transported over the network.
Clip used the standard LLC/SNAP encapsulation as defined in [RFC 1483]. Lane v1.0 uses its own encapsulation. But Lane v2.0 uses the standard encapsulation. By default, Mpoa uses LLC/SNAP encapsulation for all data flows. The default shortcut data encapsulation is also LLC/SNAP encapsulation. Mpoa also allows the optional use of the Mpoa tagged encapsulation. IP switching uses default encapsulation multiplexed over VCC= 0,15. If a flow is detected IP Switching uses default encapsulation over a reserved VCC.

5.6.2 MULTICAST

Future applications requires multicast traffic to connect multiple clients. Clip does not support this. Lane v2.0 has defined a special server for multicast traffic, this could be a special multicast server (SMS) or the Intelligent BUS (I-BUS). Mpoa v1 does not support multicast shortcuts. Multicast over the default path could be supported if Lane v2.0 is implemented in the default path. In the next release of Mpoa, v2.0, a MARS is implemented. IP switching does support multicast. With multicasting a difference must be made. Multicasting can take place on two levels, there is IP multicast and ATM multicast. There are two ways to support IP multicasting in ATM. IP multicasting can be supported by multiple ATM unicast VC's or by one multicast ATM VC. The problem there is the translation of one IP address to multiple ATM addresses. This can be done by a (special) ARP server like a SMS or MARS. Another solution can be the implementation of a separated server that takes care of all the multicast traffic.

5.6.3 RESERVED ATM VCS

Clip uses the following VCC numbers :

- 0,0 = unassigned cell
- 0,5 = UNI by SVC
- ≥0, >32 = data traffic

Lane uses the following VCCs numbers :

- 0,0 = unassigned cell
- 0,5 = ATM signalling
- 0,16 = ILMI
- ≥0,>32 = Data en LANE packets.

Mpoa uses the following VCCs numbers :

- 0,0 = unassigned cell
- 0,5 = signalling
- 0,16 = ILMI
- ≥0,>32 = Data en Lane and Mpoa packets

IP Switching uses the following VCCs numbers :

- 0,0 = unassigned cell
- 0,15 = IFMP, default encapsulation
- others = cut-through path

5.7 RESOURCE RESERVATION PROTOCOL AND ATMS QUALITY-OF-SERVICE

Quality of service is an important issue for ATM networks [ATM], in part because it is used for real-time traffic, such as audio and video. When a virtual circuit is established by means of signalling, there has to be a contract defining the service. The contract between the customer and the network has three parts, i.e. the traffic to be offered, the service agreed upon, and the compliance requirements. The first part of the contract is the traffic descriptor and characterises of the load to be offered. The second part specifics the quality of service desired by the customer and accepted by the carrier. To have concrete traffic contracts, the ATM standard defines a number of QoS parameters. These parameters are the about the speed of the traffic (*Peak Cell Rate, Sustained Cell Rate, and Minimum Cell Rate*), specification of the network characteristics (*Cell Error Ratio, Severely-Errored Cell Block Ratio, and Cell Misinsertion Rate*), the characteristics of the network measured at the receiver (*Cell Loss Ratio, Cell Transfer Delay, and Cell Delay Variation*), and the variation present in cell transmission times (*Cell Variation Delay Tolerance*). The third part of the contract tells what constitutes obeying the rules. One must notice that the user does not explicit chooses for any QoS parameter, but implicit choose for a QoS class with this information[2]. Table 1 in appendix A specify the relationship between the QoS classes. Current Internet architecture, shaped by the IP protocol, offers a simple point-to-point 'best-effort' model. This 'best-effort' service is acceptable for e-mail, world wide web (WWW), ftp, etc. In the past years several classes of distributed applications are developed. This includes remote video, multimedia conferencing, virtual reality, and picture telephony. Even present telephony demands quality in order to work acceptable over Internet. One of the most important features of these new applications is multicast traffic to more than one client. Overprovisioning by increasing the only bandwidth is not enough or cost effective. A real solution is to this is a new protocol, called Resource reServation Protocol (RSVP) as treated in chapter 3. In the migration path to ATM a match between RSVP and ATMs QoS is needed with the different protocols which support network protocols over ATM. These protocols are Clip, Lane, MPOA and IP switching.

5.7.1 RSVP CONSIDERATIONS

5.7.1.1 WHAT IS NECESSARY TO IMPLEMENT QOS IN A NETWORK

The ability to offer quality services can be subdivided into five different parts.

1) **Flow specification** - Needed to characterise a flow. A prospective is made in [RSVP1] and represented in figure 3 below.
2) **Routing** - to determine the path through the network from source to destination.
3) **Resource Reservation** - The network must make some reservations to support the QoS service.
4) **Admission Control** - Accept or deny a reservation to keep the network manageable.
5) **Packet-scheduling** - Determine which packet to send next.

[2] The next versions of the standards are more flexible; i.e. one can specify the QoS by every parameter.

0	8	15	31
Version		Max. Transmission Unit	
Token Bucket Rate		Token Bucket Size	
Max. Transmission Rate		Minimum Delay Noticed	
Max. Delay Variation		Loss Sensitivity	
Burst Loss Sensitivity		Loss interval	
Quality of Guarantee			

0	Exponent	Value

1	Well defined Constant

figure 3 : prospective flow specification

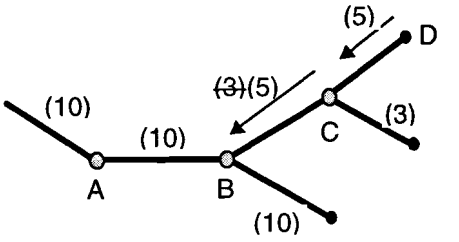


figure 4 : example reservation

RSVP is a protocol that looks at point three. In chapter 3 the way in which RSVP works is explained and is not repeated here, additional information about RSVP can be found in [RSVP2] or [RSVP3]. Before each source begins with the transmission of data-packets it must sent a 'path'-message to every destination connected, this in order to support any form of resource reservation. This 'path'-message contains the flow specifications of the data source. If the 'path'-message arrives at a destination a reservation can be made by this destination. This is done by sending a 'reservation-message' back into the network. This 'reservation'-message must follow the same route as the 'path'-message only backwards. This reservation follows the backwards path until it meets a point where already a reservation of the same flow is made, higher than this reservation. This is explained in figure 4, where in point D a reservation is made for e.g. a delay or bandwidth. This reservation is send back till point B, because from this point on there already exists a higher reservation for the flow towards the source.

5.7.1.2 RSVP SPECIFIC PARTS

One of the most important features of ATM is its ability to deliver a point-to-point or point-to-multipoint connection with a specific **QoS**. Because of this build in feature is self-evident that RSVP wants to take advantage of this ability. One important point in the acceptance of ATM QoS in RSVP is the integration of RSVP signalling and ATM signalling. This consists of two parts.

1) **QoS parameter translation** - the mapping of RSVP QoS parameters to the correct ATM QoS parameters.
2) **VC management** - This concerns the questions, " How many VC are necessary' and "What kind of traffic is send over which VC".

The translation of QoS parameters between RSVP and ATM is not the hardest problem, the concentration goes out to point two, the VC management. This part can also be divided in to three parts.

1) **Reservation style** - type of reservation
2) **Dynamically changing a reservation** - 'on the 'flow' changing the QoS.
3) **Heterogenic reservations within a multicast environment** - different QoS in one multicast flow.

5.7.1.3 DIFFERENT TYPE OF RESERVATION STYLES WITHIN RSVP

The reservation style stipulates the amount of RSVP flows necessary within a multicast group, and determines which source belong to a flow. A reservation can be 'distinct' or 'shared', this depends on the number of sources going within the flow. A reservation style can also have a 'explicit' or a 'wildcard' sender selection, this depends on the fact that one can select a specific source or the fact that one can select every source. Each reservation style is suitable for a particular type of traffic. The different reservation style are arranged in figure 5. It is not clear where to use each reservation style.

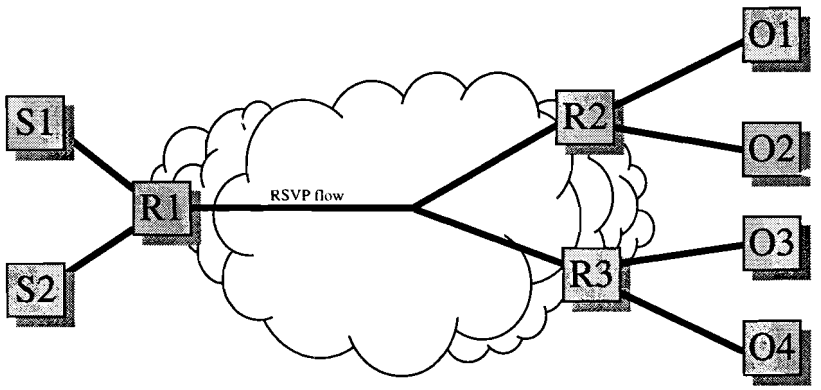
	Reservation	
Sender Selection	Distinct	Shared
Explicit	Fixed Filer Style (FF)	Shared-Explicit Style (SE)
Wildcard	(not defined)	Wildcard-filter style (WF)

	Reservation	
Sender Selection	Distinct	Shared
Explicit	video conferencing	Audio conferencing
Wildcard	-	Audio conferencing

figure 5 : Reservation styles and examples

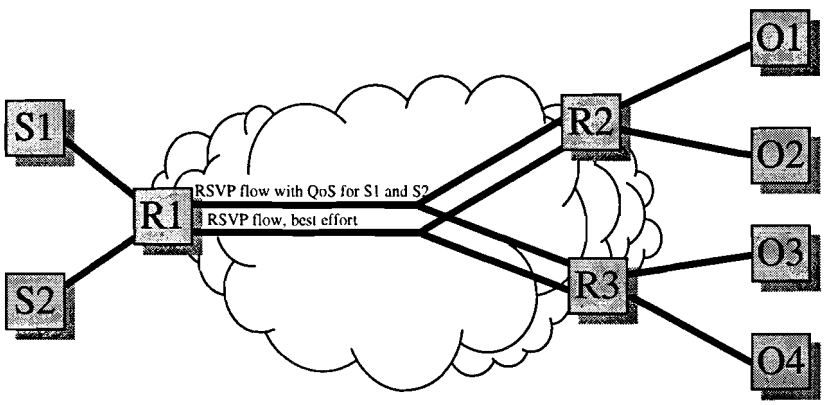
5.7.1.4 NUMBER OF RSVP FLOWS IN AN ATM NETWORK BASED ON A RSVP STYLE

This chapter has a close binding with the next chapter 5.7.1.3, where the RSVP are translated to ATM VCs.



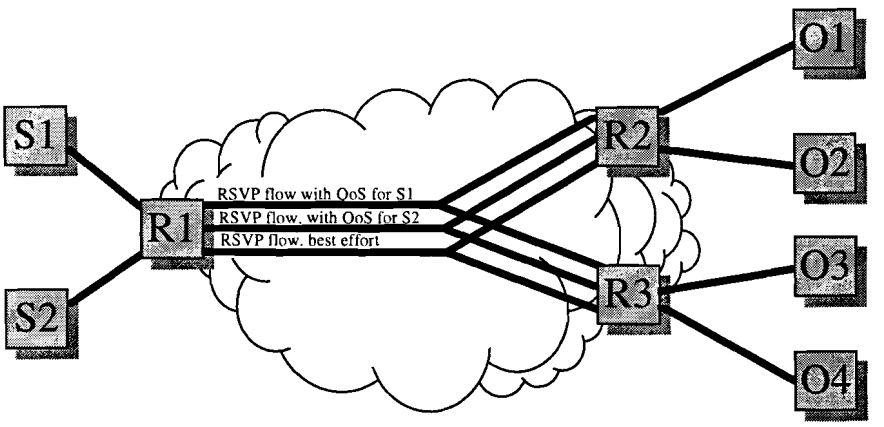
A reservation with a WF style requires in an ATM-network one RSVP flow, because every source is part of the same reservation. The QoS of this RSVP flow is that of the highest requested one by the receivers.

figure 6: WF reservation within ATM network



A reservation with a SE style requires two RSVP flows in an ATM-network. One flow that will deliver the highest requested Qos for that flow. The other flow is necessary to deliver the 'best-effort' guarantees to those receiver who did not ask for a specific QoS.

figure 7 : SE reservation within ATM network



A reservation with a FF style requires n+1 RSVP flows within an ATM network. Every present source requires one RSVP flow, counting up till n. And one flow to deliver the 'best-effort' guarantees to those receiver who did not ask for a specific QoS.

figure 8 : FF reservation within ATM network

5.7.1.5 TRANSLATION OF RSVP STYLES INTO ATM VCS[3]

There are different approaches to translate the RSVP flow of chapter 5.7.1.4 into ATM VCs. Here, two of these approaches will be discussed and a third one is short mentioned.

1) **One VC per flow model** - Each RSVP flow is directly translated into a ATM VCC. The problem with this model is that each receiver demands a different reservation. This can be solved by equipping the VCC with a QoS matching the highest requested reservation.
2) **More VCs per flow model** - Each flow can be translated into more different ATM VCCs. One specific flow can be translated to different VCCs with it's one QoS. The advantage of this is that one can deal with the different reservation of the receivers.
   a) Two ATM VCs per RSVP flow
      Within each a flow a 'best-effort' VCC is established, because there are receivers who do not request a reservation. For the receivers who do request a reservation another VC with QoS support is established. Within this VC the highest requested reservation QoS can be chosen. If each of the receivers makes a reservation, the 'best-effort' VC is no longer necessary.
   b) More than two ATM VCCs per RSVP flow
      Instead of specify one VCC for the highest reservation, one can translate the reservation flow into more VCCs. Each of these flows carries the requested reservation. The disadvantage of this method is that more copies of the same data packet are transported over the same link, and this method asks for more network resources. One the other hand it adds more manageability to the network policies.
3) **grouping RSVP flows in a ATM VCC.** Only mentioned.

Examples

[3] In these and following chapters frequently the term 'VC with a specific QoS' will be mentioned. This specific QoS means those characteristics corresponding with requested reservation.

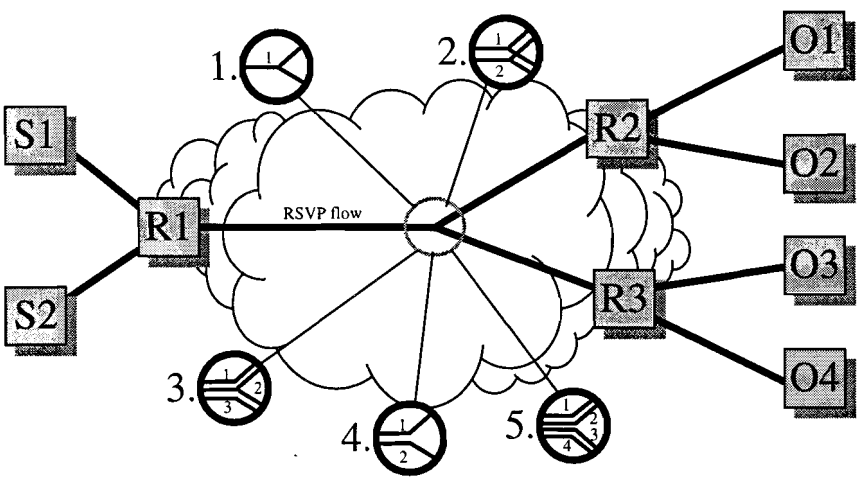


figure 9 : WF reservation to ATM VCs translation

The translation of a WF reservation can take place in five ways. Case 1 to 3 can only be implemented if multicast is supported.

1) One VC per flow
2) Two VCs per flow, having one 'best-effort' and one with QoS.
3) More VCs per flow, whereof one 'best-effort' (2) and the others with the requested QoS reservation (1,3).
4) Same as 1), but without the support of p-t-m[4] VCs.
5) Same as 2), but without the support of p.t.m[1] VCs.

The translation of a SE reservation style can take place in two ways.

1) One VC per flow.
2) Same as 1), but without the support of p-t-m[1]

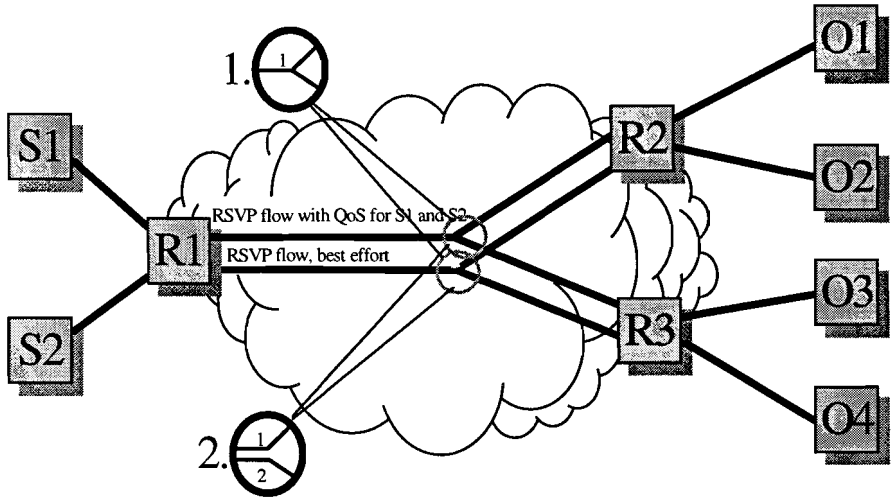


figure 10 : SE reservation to ATM VC translation

The FF reservation style can, like the SE style, be translated in two ways

1) One VC per flow
2) Same as 1), but without the support of p-t-m[1]

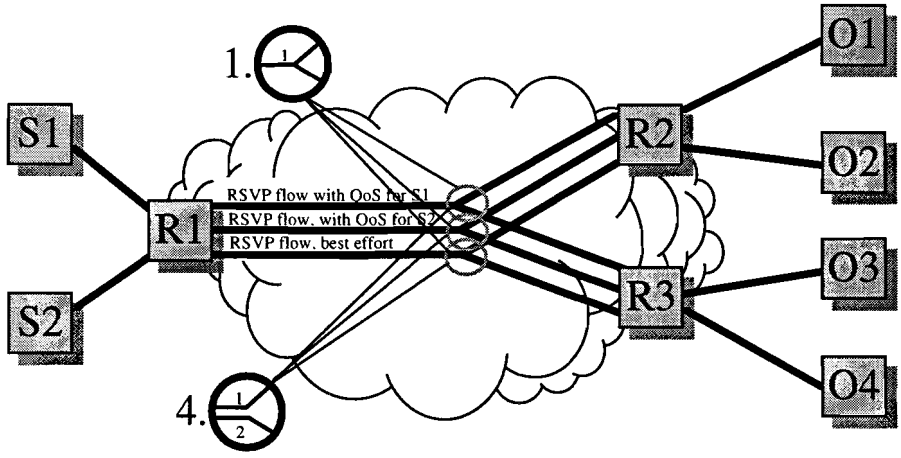


figure 11 : FF reservation to ATM VC transation

Dynamic QoS

One of the issues that

[4] Point-to-multipoint

can give a major problem is the support for dynamic QoS. This deals with the fact that any receiver may and can change his reservation at any time. This is complete supported by RSVP, but within ATM dynamically changing the QoS is not supported. Even the new version 4.0 of the ATM UNI signalling protocol specifications [UNI4.0] does not do this. As mentioned before RSVP does support dynamic QoS, in the way that the requested QoS can be change at any moment. Of course with this the following question will arise, "Why to change a reservation request?". For this there are several answers possible :

1) The present QoS is not acceptable anymore.
2) The sender changes it's traffic-specifications, by sending a new 'path'-message.
3) A new sender, with different (higher) 'path'-parameters can be used.
4) A receiver can change it's reservation by increasing the old reservation.

The disadvantage of ATMs, is its connection-oriented nature. That is why the old VC must be teared down before and set-up a new one with the right QoS. This is a time-critical action, which adds more delay to the network. One solution can be to create the VC first with the new QoS parameters and afterwards tear down the old one. The delay stays the same but the advantage is that if the network can not deliver the requested QoS because of the lack of resources, the old one still exists. Another solution can be the usage of timers and collecting all the new reservation requests belonging to one flow. When the timer alarms, all the reservations are handled at once.

5.7.1.6 THE TRANSPORT OF RSVP PACKETS

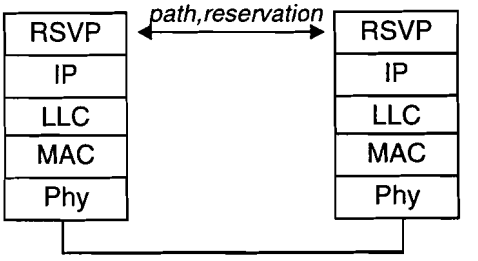


figure 12: Position of RSVP in the ISO-OSI model

In the Internet, RSVP packets are send via the IP protocol (version 4 and 6), see figure 12. Within RSVP there are two kinds of packets, the 'path'-message and the 'reservation'-message. The 'path'-message is send from sender to receiver and can use multicasting. The 'reservation'-messages can not use multicasting, because they are receiver explicit and must use the same path, but backwards, as the 'path'-message. There are four ways the RSVP packets can be sent in ATM.

1) Use a data VC - this has as advantage that no additional VCs are used, but has as disadvantage that messages can be dropped.
2) separate VC per RSVP session - the advantage of this is that it is simple and clear, but has as disadvantage the use of additional VCs and therefor waste bandwidth.
3) separate point-to-multipoint VC with multiplexed RSVP sessions - A separate point-to-multipoint VC from ingress router to every egress router. The advantage is that it uses less VC's as separate VC per RSVP session (see point 2).
4) more point-to-point VCs multiplexed over the RSVP sessions - same as 3), but with additional point-to-point VCs per multicast VC for the 'reservation'-messages.
5) separated VC for RSVP signalling - Make use of one of the reserved ATM VCs, like ATM signalling does.

5.7.1.7 ADDITIONAL COMMENTS ON RSVP

The following points, which have not been addressed earlier must be taken in consideration.

1) A sender S1 can deliver 10Mb/s. If a receiver asks for a reservation of 4Mb/s, the question is, "what is done with the remaining 6Mb/s?". Is this remaining part discarded or delivered as 'best-effort'.

2) How is decided which data packets belong to a reservation and which are part of the 'best-effort' part.
3) ATM is source-driven and RSVP is destination-driven. This means that a RSVP reservation is build up starting with the destination and within ATM the QoS of a connection is set-up at the same time as the connection itself.

5.7.2 RSVP AND CLIP

The support of RSVP by Clip is the translation done in section 5.7.1.5. In the case of Clip there aren't many problems. Most of the problems are partly solved by the support of unicast traffic. The use of Clip in combination with permanent VC's not recommended, because with that it is not possibility to make reservations. The use of switched VCs does support making reservations.

5.7.3 RSVP AND LANE

Lane is extremely well capable for intra-networks. For communication between these networks, routers are required. For the routers and the Lane Clients it is necessary that they support RSVP. To avoid problems with the use of RSVP in combination with Lane the following assumptions are made.

LAN emulation service must support RSVP *(1)*

If the Lane services, including Lane Clients, do not support RSVP, it is not possible to offer QoS at all. The deployment of RSVP in a Lane environment gives an additional problem as discussed in section 5.7.3.1.

A data direct connection is already set-up by Lane. *(2)*

It is hard to make a reservation before a data direct connection is set-up. This is because some of the data may be sent through the BUS, however that connection cannot deliver any QoS besides the 'best-effort' service.

Multicast is supported by LANE *(3)*

This concerns only Lane v2. Lane v1 does not support multicast at all. The present of multicasting is of great use for RSVP and has it's benefits above unicast connections.

One sender per flow *(4)*

The use of multiple senders per flow makes the implementation of RSVP more difficult than it is. If the need of multiple senders is necessary, it is recommendable to use them as separate RSVP flows. Thus if necessary equip both flows with a 'best-effort' flow.

5.7.3.1 RSVP SUPPORT BY LANE

As mentioned before there are some difficulties supporting RSVP in a Lane environment. These difficulties are caused be the Lane feature to hide the ATM functionality. So the ATM QoS is unreachable for the Lane Client. This can be solved as follows :

1) Adapt Lane - Implement an additional Lane/RSVP functionality to support the usage of ATMs QoS.
2) Adapt RSVP in combination of Lane - One could think about providing ATM with RSVP information partly bypassing Lane.

The first solution has as major disadvantage that the contents of every packet should be checked, so that a RSVP packet can be intercept, this adds additional latency. The adaptation of RSVP has as major disadvantage that there could appear different versions of the RSVP protocol, one with and one without Lane support. Also one advantage of Lane, the one of hiding ATM for higher layers, is discarded. Another problem is the absence of layer three in a Lane Client, in the case of a bridge. Because of this solution one has to be preferred above solution two.

5.7.3.2 SENDING RSVP PACKETS WITH LANE.

The RSVP packets can be sent in Lane in two different ways :

1) Each RSVP packet can use a separate connection. With the present of the Lane service this connection already exists, through the LES or through the BUS. This makes the transportation of these packets efficient and fast. The extra trouble of an extra data connection is not necessary because the overhead of RSVP packets is not worth mentioning, they are short and not very frequently.
2) On account of assumption two, Lane has already set-up a data direct connection before a reservation has been made. This data direct connection be used to transport the RSVP messages. This connection already offers the 'best-effort' service. The usage of this connection depends on the type. If it's a multicast VC it could be used to transport path-messages, otherwise they must be sent separated.

The best method is a combination of the two. The path messages could be sent by using the data-direct connection, and the 'resv'-messages should be sent by using the BUS.

5.7.3.3 VC MANAGEMENT IN LANE

This chapter does not deal with all the reservation again, but gives a description of every implementation from figure 9 till figure 11 in combination with Lane.

When figure 9 is used in combination with Lane, the routers R1, R2, and R3 are equipped with a LEC. As assumed already a point-to-multipoint data-direct connection is set-up to transport the best-effort delivery. After receiving a receiver can make a reservation, a reservation messages is send back the sender. This request is also send from R2 to R1. When the LEC of R1 recognises this reservation request, the following choices can be made :

1) Changing the 'best-effort' VC to each receiver into a VC with the QoS of the requested reservation.
2) Keep the 'best-effort' VC, but tear down the part issuing R2 and replace this part with a VC with the requested reservation.
3) If already a VC is set-up with a reservation beside the 'best-effort' VC then
   a) used this one, by adding R2 as multiparty, and if necessary upgrade it.
   b) Set-up another VC beside this one, with the requested reservation.

The same solution can be used within figure 10 and figure 11 as long as one sender is used.

5.7.3.4 DYNAMICALLY QOS IN LANE

When changing the QoS 'on the flow', a new VC must be set-up with a different QoS, because ATM does not support on the flow changes. The whole end-to-end connection must be teared down and set-up again.

5.7.4 RSVP AND MPOA

One of the great advantages of Mpoa is creating a shortcut, and uses this to bypass routers. To avoid problems the following assumption are made related to the use of RSVP in combination with Mpoa.

Mpoa must support RSVP ***(1)***

If Mpoa, including the Mpoa clients don't support RSVP, it is not possible to offer QoS at all. The deployment of RSVP in a Mpoa's default path gives an additional problem, that is why assumption (2) is made.

Mpoa already has set-up a 'best-effort' shortcut ***(2)***

It is hard to make a reservation before a shortcut is set-up. This is because the data is send by using the default path. This can lead to double reservations for a receiver. Any reservation set-up using the default path must be tear down when setting up a shortcut. If not enough traffic is send to trigger the flow detection mechanism, the Mpoa Server can trigger the Mpoa Client.

Multicast is supported by Mpoa ***(3)***

This concerns the current version of Mpoa, but future versions will support. But the present assumption of supporting multicast is of great use for RSVP and has it's benefits above unicast connections. So it is assumed but not present currently.

The entire path uses ATM ***(4)***

Within an environment with no hosts, like a backbone, connected to the ATM network, there is no direct benefit of using Mpoa. Lane or Clip could be used instead. However there is a benefit when ATM is implemented completely to the desktop or receivers and this assumption is made here. The use of Mpoa in this case can create the benefit of bypassing routers.

5.7.4.1 SUPPORT OF RSVP BY MPOA

Mpoa is a protocol that is operating on layer two and three, and it's using the Lane technique to support network protocols in the default path. Because Mpoa uses layer three, it does not have the same problem as Lane. Mpoa works with 'flows', defined as packets designated for the same network destination. The direct use of Mpoa direct under the IP stack, makes it easy to support RSVP. This situation is clarified in figure 13.

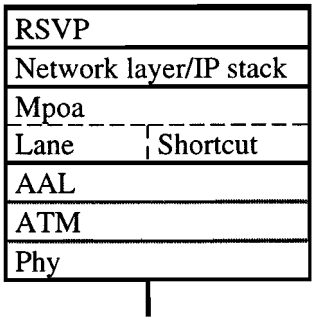


figure 13: Mpoa stack with RSVP

5.7.4.2 SENDING RSVP PACKETS BY MPOA

The assumption that a shortcut is already made, makes it a lot easier to send RSVP packets. RSVP packets can be sent together with the packets used for the target resolution process. If the flow detection mechanism in the Mpoa Client detects that a shortcut should be established the TLV-fields in *'Mpoa resolution'*-packets could be used to transport the 'path'-messages from Mpoa Client to Mpoa Server. The Mpoa Server translates the resolution packet into NHRP packets, hereby copying the TLV-fields. The NRHP packets are translated by the egress Mpoa Server into Mpoa impositions packets. Again the TLV-fields are copied. In the reply from the Mpoa impositions request, NRHP requests, and Mpoa resolution request the same TLV-fields can be used to transport the 'Reservation'-requests. Depending on the kind of reservation a shortcut can be set-up with the correct QoS. One must take into account that before setting up this shortcut a new Mpoa resolution request must be sent to the egress Mpoa Client, because a VCC with another QoS is requested. This must be affirmed be the egress Mpoa Client.

5.7.4.3 VC MANAGEMENT IN MPOA

Like the discussing in section 5.7.3.3 about the VC management in Lane, the different reservation styles aren't discusses, but a description of every implementation from figure 9 till figure 11 in combination with Mpoa is given. When an ATM network is going to be used in these figures, the routers R1, R2, and R3 are replaced by Mpoa Servers, and the senders and receivers are Mpoa Clients. When a shortcut already exists and the QoS equals to the 'best-effort' reservation and a reservation is made, the following actions can be made :

1) Changing the 'best-effort' shortcut to every Mpoa Client into a shortcut with the requested QoS.
2) Keep the 'best-effort' VC, but tear down the part issuing R2 and replace this part with a VC with the requested reservation.
3) If already a shortcut is set-up with a reservation beside the 'best-effort' shortcut then
   a) used this one, to by adding R2 to it as multiparty. And if necessary upgrade it.
   b) Set-up another shortcut beside this one, with the requested reservation.

The flows of the other two figures have the same solutions, only the situation with two senders can't be resolved by Mpoa yet. This can be solved by dealing the two senders as different one's.

5.7.4.4 DYNAMICALLY CHANGING THE QOS WITHIN MPOA

When a shortcut is made and equipped with the QoS of a Reservation, and receivers want to change this, a shortcut must change his QoS on the flow. This is not supported by ATM. Mpoa can solve this by tearing the old shortcut and request a new one. The danger of asking a new shortcut can be that the ATM network doesn't accept the new shortcut. This has as result that the old one has to be restored, resulting in the reservation with the old QoS.

5.7.5 RSVP AND IP SWITCHING

One of the features of IP switching is changing from routing to switching within the same device. To avoid problems the following assumptions are made :

A Cut-through path is already made ***(1)***

Supporting RSVP in the routed default path of IP switching does not offer correct reservations. Because all the traffic is send by using the same connection, you never know the amount of traffic send over this connection. It's easier to deliver QoS when a separate connection can be used. This connection supports the 'best-effort' reservation.

IP Switching supports RSVP *(2)*

When IP switching doesn't support RSVP, it is impossible to deliver QoS into the network.

5.7.5.1 SUPPORT OF RSVP BY IP SWITCHING

The support of RSVP by IP switching is the easiest comparing to the other techniques. This is because the same path, as in legacy LANs is used. The only difference is that the router is changed into a ATM switch and the change of transport mechanism into ATM. The entities where RSVP is supported remains on the same location. And it is still possible to use RSVP in the same component, because the don't have to change into a LEC or Mpoa server.

5.7.5.2 SENDING RSVP PACKETS BY IP SWITCHING

The 'path' and 'resv'-message are send the same way as in legacy networks. The packets are encapsulated into ATM cells to transport them by using the default encapsulation.

5.7.5.3 VC MANAGEMENT IN IP SWITCHING

Again, only some descriptions are given of some examples of figure 9 till figure 11.

When changing the ATM network of figure 9 into an IP switching network, the routers R1, R2, and R3 are replace by an IP switch. The receivers and senders are equipped with an IP switch interface. When a cut-through path is set-up every client of the flow can uses the 'best-effort' service for that flow. When a client makes a new reservation an IP switch, e.g. R2, the following actions can be made:

1. Changing the QoS of the cut-through path between it's neighbours, as shown in
2. figure 14.
3. Maintain the 'best-effort' cut-through path to other participants, and reserve a new cut-through path for the requested reservation.
4. If already a cut-through path is reserved for a reservation beside the 'best-effort' path
   a) then use this reservation and if necessary upgrade it.
   b) reserve a new cut-through path with the requested reservation.

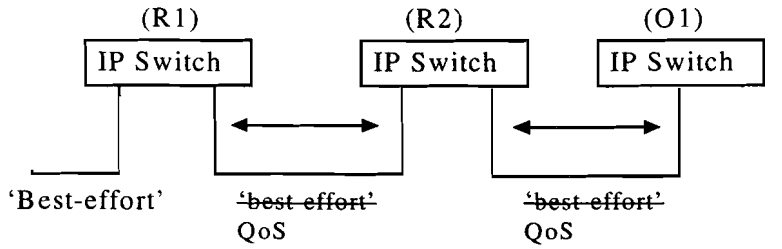


figure 14 : Change a 'best-effort' reservation into a requested QoS

5.7.5.4 DYNAMICALLY CHANGING THE QOS WITHIN IP SWITCHING

When a cut-through path is reserved that is equipped with the requested QoS and a new reservation is made to change the QoS on the flow only some communication is needed between some adjacent IP switches to change the QoS. If the change is accepted then the QoS is being updated. A IP switch flow is not a end-to-end flow like with Mpoa, but an IP switch flow is composed of several cut-through connections. When a change of a reservation is made, it is possible that some cut-through connection will accept the QoS. You'll get a flow with different end-to-end QoS. This isn't only an IP switching problem, but also appears in traditional IP networks.

5.8 EXPERIENCE IN THE FIELD AND AVAILABLE PRODUCTS

This sections discussed the experience that is available by KPN research on Clip, Lane, Mpoa, and IP switching

5.8.1 EXPERIENCE IN THE FIELD

The following experience is gained at KPN research :

Clip

In August 1995 Marc v.d. Bergh, at that time a student of the Eindhoven Technical University, has done some research on the performance of TCP/IP over ATM, hereby using the "(Classical) IP over ATM" approach with permanent connections. His research can be read in his master thesis R&D-SV-95-692 [Marc].

Lane v1.0 and v2.0

Approximately one year later, some experiments where done with a Lane emulation pilot using version 1.0. Later that year I did some field work on the performance of Lane together with a student of the Nijmegen University. The functionality of Lane version 2.0 is discussed by that other student in [Sloots], but no field experience is gained.

Mpoa

Because Mpoa is a fairly new protocol, that isn't developed completely, there aren't much products to test. Therefore there is not many experience in the field. Only theoretical research is done and described in this document.

IP switching

From January 1997 until now a student of the Delft Technical University is looking at the functionality of IP switching. This includes some technical research on IP switching products in July this year.

5.8.2 PRODUCTS AND IMPLEMENTATIONS

The products available :

Clip is the most used technology to transport IP over ATM, because it is the first that is fully developed and it is the most simple one to implement.

The **Lane** specification version 1.0 is the next one being developed and several products and implementations are available, by CISCO and FORE. The Lane v2.0 specification is still in development and no products with implementations are available.

The **Mpoa** standards is also still evolving but several products are yet development by NEWBRIDGE. The **IP switching** world is much faster, because the standard is developed later and the product are available earlier. These products are developed by IPSILON.

5.9 SUMMARY ON WEAK AND STRONG POINTS

This section discusses the weak and strong point of IP networks and some of the underlying legacy techniques and the techniques used to transport network protocols over ATM.

5.9.1 LEGACY LANS

Techniques for transporting network layer protocols are : Ethernet, token ring, and FDDI.

Ethernet (CSMA/CD 802.3)
This technique is also named 10BASE-T Ethernet. Summarised Ethernet has the following pro's. Beside it is the most used technique, it is a simple protocol and when there is less network traffic, the delays are low. The disadvantages are that it uses an analogue technique. Therefor it is not suitable for real-time applications and collisions are a great problem on heavy duty networks. Furthermore, no priorities are possible and the bandwidth is 10mb/s.

Switched 100BASE-T (Fast Ethernet)
Like Ethernet and Token Ring, switched 100BASE-T, or Fast Ethernet, is based on a shared-medium approach. A major advantage of 100BASE-T is that it functions identically to 10BASE-T, but operates at 10 times the speed at a nominal cost increase. The advantages are that it uses a well-understood technology and more bandwidth is available. The disadvantages are the same as 10BASE-T Ethernet.

Token ring(802.5)
The advantages of a token ring network are, that simple management and priorities are possible, efficient by high throughput but not fair. The disadvantages are : no quality guarantees, inefficient with low loads and not suitable for real-time applications. Furthermore it has a single point of failure at the "central monitor". This monitor takes care of the token management.

FDDI
FDDI uses a similar technique as token ring. The advantages are the higher bandwidth and the disadvantages are the few competing products and thus leaving the costs rather high, and no industry standard for full-duplex operation is specified.

Network protocols
Only the most used network protocol, the Internet Protocol IP is discussed. This protocol has the features that, it is a connectionless network, it uses different packet sizes, the network takes care of the transport and deliverance of the traffic. One of the major advantages is that it is a simple and flexible network. But the disadvantages are the low capacity of the used lower layer techniques, it's 'best-effort' nature, and need of routers between networks.

5.9.2 TECHNIQUES FOR TRANSPORTING IP OVER ATM

Summarising the previous chapters, the following gives a overview of the advantages and disadvantages of Clip, Lane, Mpoa, and IP switching.

CLIP

Clip has the following advantages :
1) Completely compatible with IPv4.
2) Large MTU size.
3) No changes in higher layers.
4) Enables simple integration with IP-based service and ATM services.
5) In the future QoS is supported in combination with RSVP.

The disadvantages are :
1) Only for IP.
2) Bridging is not possible.
3) Needs routers between two LISses.
4) The used routers are a bottleneck.
5) QoS uses the traditional routing protocols and offers only 'best-effort' delivery.
6) No reuse of existing legacy LAN infrastructures.
7) Multicast and unicast is not supported.
8) No default path exists before a connection.

LANEv1

Lane v1 has the following advantages,
1) High bandwidth by using ATM.
2) No change in higher layers.
3) Capable of reusing legacy LAN equipment.
4) ATM is invisible.
5) Makes a migration path LANs and ATM

The disadvantages are :
1) Has single point of failure in the BUS/LES/LECS.
2) BUS is the bottleneck of Lane.
3) Not efficient and complex.
4) Needs a lot of data connections.
5) Has a high initialisation time.
6) Not easy to scale.
7) Needs routers between elans.
8) Requires double address resolution.
9) ATM's QoS is invisible.
10) Can't resolve translation problems between technologies like Ethernet, Token Ring and FDDI. A bridge/router is required to handle conversion between these technologies.

LANEv2

Lane v2.0 has the following advantages :
1) No single point of failure.
2) More scaleability.
3) Better performance of LES-functionality, by locating by the use of multiple LESs, so a LEC can be closer by a LES.
4) Better performance of BUS functionality.
5) Makes a migration path LANs and ATM

The disadvantages are :
1) Still needs routers to communicates between elans.
2) The routers can still be a bottleneck.

MPOA

The advantages of Mpoa are :
1) Allows efficient communication between subnetworks.
2) Avoiding router bottlenecks.
3) Improves manageability.
4) Improves scaleablity.
5) Enables multicast and broadcast on layer three.
6) Flexibility.
7) Flow detection on the hosts.
8) Makes a migration path LANs and ATM.

The disadvantages are :
1) Very complex protocol.
2) Only available for IPv4 and IPX.
3) Host stack must be changed.
4) Default path still has lots of delay.
5) Lane can be a bottleneck.

IP Switching

The advantages of IP switching are :
1) Only for IP.
2) More simple by
   a) not using the ATM signalling.
   b) ATM hardware is cheap.

The disadvantage are :
1) It uses the routed path through the network.
2) Flow detection takes place in every IP switch.
3) Offers QoS by hop, because it still is a hop-by-hop protocol.
4) Creates flat networks.
5) Uses a lot of chit-chat between IP switches reduces the available bandwidth.
6) Not available in a bridged network, by the lack of a IP-address in a bridge.

5.10 IP- EN ATM FEATURES

This chapter gives the present IP and ATM features in Mpoa, Clip, Lane, and IP switching. Based on the previous sections.

IP en ATM-features	Mpoa	Clip	IP switching	Lane version 1.0	version 2.0
Network support					
IPv4 Unicast	yes	yes	yes	yes	yes
IPv4 Multicast	MPOA v2	-	yes	-	is possible
IPv4 broadcast	MPOAv2	-	-	yes	yes
Support IP v6 (IPng)	-	-	-	-	-
IPv4 router-protocols ('OSPF','RIP', etc.)	RIP, OSPF	standard router-protocol	RIP, OSPF, CIDR, DVMRP	standard router-protocols	
IPv4 Router functionality	in Mpoa Server	in traditional routers	Router discovery, IGMP, ICMP, ARP	in traditional routers	
future RSVP support	yes	yes	yes	difficult	
Connection					
Connection type	connection-oriented	connection-oriented	partly Connectionless	connection-oriented	
Multi-acces	yes	yes	yes	yes	yes
Permanent/Switched Connection	SVC	PVC/SVC	SVC	SVC	SVC
ATM unicast	yes	yes	yes	yes	yes
ATM multicast	MPOA v2	-	yes	-	SMS/IBUS
ATM direction	Uni/bi	Uni/bi	Uni/?	bi	uni/bi
ATM addressing	DCC, ICD, E.164	E.164[5]	?	See [sloots][6]	[sloots]
QoS	**MPOA v2** prior QoS negotiation and uniform QoS for all VC's	only possible at the ATM level	Dynamic QoS and heterogeneous QoS	only possible at the ATM level	
ATM routing	Uni, PNNI, I-PNNI	Uni	proprietary	Uni	Uni
ATM signalling	Uni 3.0, 3.1, 4.0	Uni	not applicable	Uni	Uni
Security	yes	-	-	-	-
Hide underlying structure	no	no	Does not use AAL layers	yes	yes
Flow detection	at the edge	no	at every IP switch	no	no
ATM Transfer Capabilities	UBR	UBR	'best-effort'	UBR	UBR, VBR
Encapsulation	LLC/SNAP	LLC/SNAP	LLC/SNAP	LANE encap.	LLC/SNAP
Reuse of equipment					
Legacy host	yes	yes	yes	yes	yes
legacy routers	yes	yes	yes	yes	yes
'old' ATM switches	yes	yes	yes	yes	yes
Kind of new equipment	NIC-card (ATM-attached) edge device (bridge) MPOA servers	NIC-card + software	IP switch gateway IP switches or CLIP	NIC-card LANE/ATM switches	NIC-card LANE/ATM switches
Availability today					
Standard	June. '97 Straw ballot	Jan. '94 (RFC1577)	may '96	Jan. '95	½ '97
Equipment	begin '97 (vivid)	yes	yes	yes	-

Table 2 : comparison between Clip, Lane, Mpoa, and IP switching

[5] source RFC 1577
[6] af-lane-0021-01 doesn't say anything about it, only 3COM uses the ICD format.

6. WHEN IS MPOA OF INTEREST

In this chapter several examples networks are given. For each network solution one or more examples are given. The solutions discussed are Clip, LAN emulation, Multi-protocol over ATM, and IP switching. In the previous chapter a comparison is made between Clip, Lane, MPOA, and IP Switching. Chapter 6 gives of each of these techniques one or more network examples, supposing that products are available. With these implementations some examples are given to indicate when and where to use a technique in the most efficient way. Together with these implementations the strong points of the implementation are mentioned. Giving network examples is the best way to show, how to use a technique. When you're starting with one specific network and then try to implement every technique is a hard and unrewarding job. This because the best implementation of a technique depends on the kind of component, size of the network, etc.

6.1 THE USAGE OF CLIP

The strongest point of Clip is it's simplicity and the support of great MTU-sizes. The weak points are the need of routers, the leak of bridging support, the support of only the IP network protocol, and the poor management function. The best situation to use Clip is as backbone and in a small intra-network.

In figure 1 an example is given of an implementation of Clip in a backbone. The ATM-switch is equipped with the IP-ATM-ARP server to resolve the ATM-addresses. The Hop-to-Hop structure of IP can be used here to effectively transport the IP packets. This is done by simply setting up one ATM connection between each router. Each packet can be send very efficient by using the LLC/SNAP-encapsulation of Clip. The high bandwidth of ATM speeds up the transport of packets.

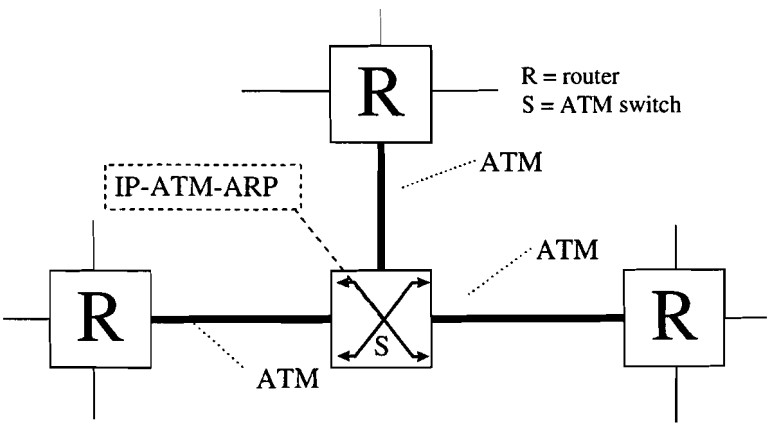


figure 1 : Clip in a backbone

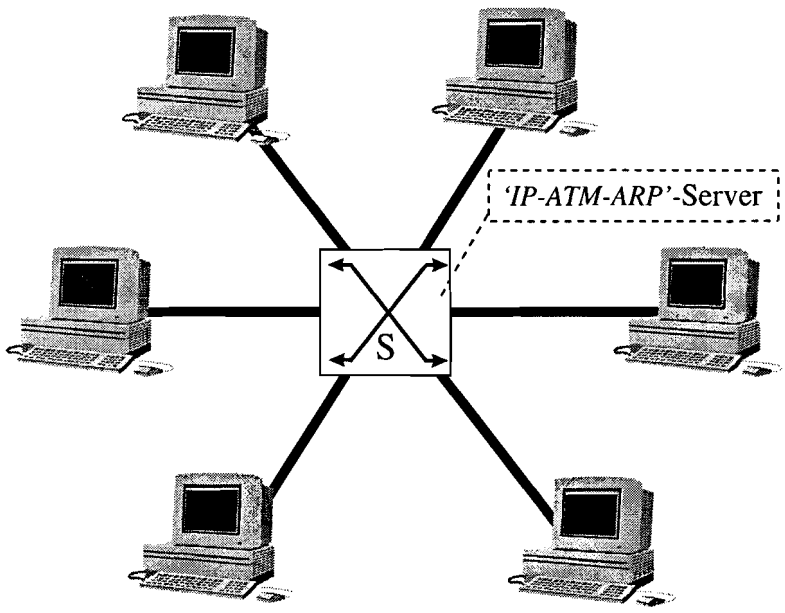


figure 2 : Clip in a intra-network

In figure 2 an example is given of an implementation of Clip in a intra-network. The advantage of such a implementation above the use of Lane is the simplicity of the usage of Clip. As long as the intra network is not too big, Clip can deal with it, otherwise Lane must be used to take care of the management of the network. If no link is needed between two networks, by means of a bridge, Clip can be used, but if bridges are present, Lane must be used. Because Clip does not support bridging by the lack of an IP-address.

Conclusion

Clip can be used in networks because of it's simplicity and speed. But if the network consist of many clients and bridges Clip may not provide a satisfying performance. If a network must be manageable, clip is also not suitable, because Clip does not delivery easy management.

6.2 THE USAGE OF LANE V2

The strong point of Lane is it's great management capability. The weak point is the need of routers between elans. The best situation is to use Lane in intra-networks .

Lane can be used to connect LANs that already are connected to other networks by bridges. This implementation is shown in figure 3. The advantage of Lane is to create virtual LANs. This is done by assigning clients of the same LAN to different elans. With this it is possible to great virtual LANs, where each computer is independent of it's location. The LECS in the Lane environment takes care of the management of the network.

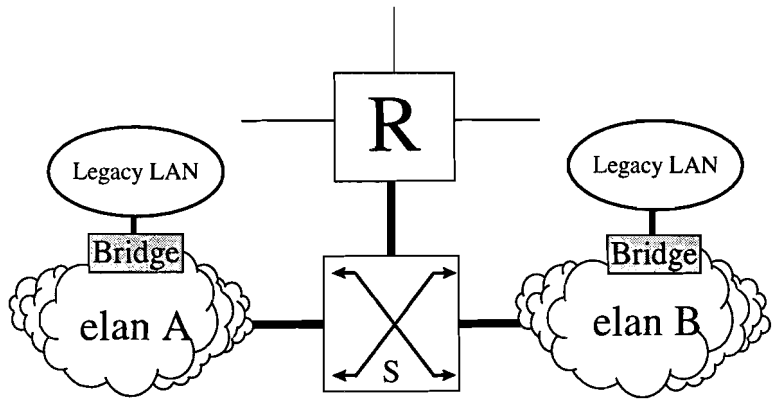


figure 3 : Implementation of Lane

An example of a virtual network is given in figure 4. Two LAN networks

are given. LAN A exists of two bridges, B1 and B2. Also a number of N clients are connected in LAN A, numbered A1 till An. LAN B has the same configuration as LAN A. The components of LAN A and LAN B can be distributed as shown in figure 5. Elan A consists of two bridges one from LAN A and one from LAN B. The clients from LAN A an B are also distributed over the elans. Elan A holds clients A1 till Ai and Bi till Bn. Elan B holds the remaining clients and bridges. In this situation clients that are communicating a lot with each other, even if they are in different LANs, can be put together in one elan. Hereby decreasing the communication overhead of normal LAN networks.

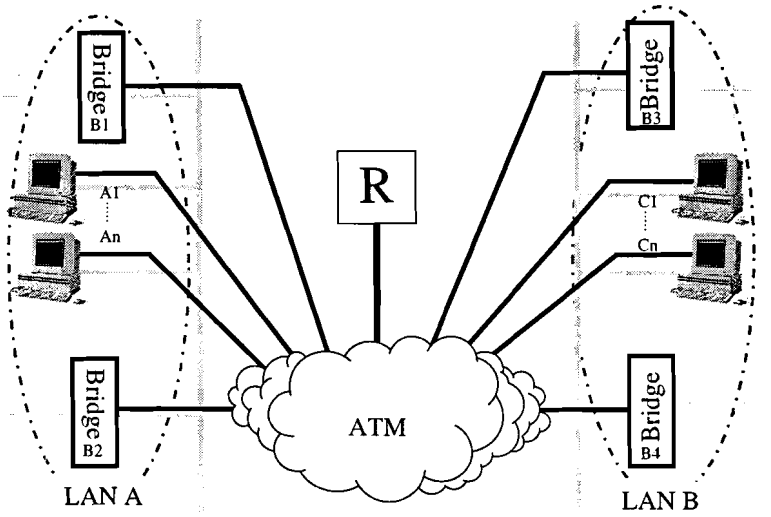


figure 4 : Lane as a virtual LAN

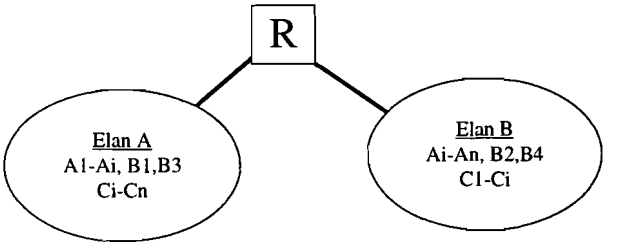


figure 5 : elan arrangement

Conclusion

Lane is suitable to be used in a backbone, in a building or a campus when a lot of clients need to be connected. Further Lane also is suitable in big intra-networks and in WAN environments where distances between LANs must by increased by implementing ATM. As mentioned before, Lane is suitable in intra-networks even if one, in exception two, routers are present in the data-path. The routers can be avoided partly by assigning the right clients to the right elan. But routers are still needed between elans. If to many routers are needed Mpoa or IP switching can be used.

6.3 THE USAGE OF MPOA

There are several situations where Mpoa can be used, these are :

1) Networks with ATM support.
2) Networks with frequent changing topologies
3) Future networks where QoS is needed.
4) Networks with different network-protocols
5) Networks where router bottleneck exist.

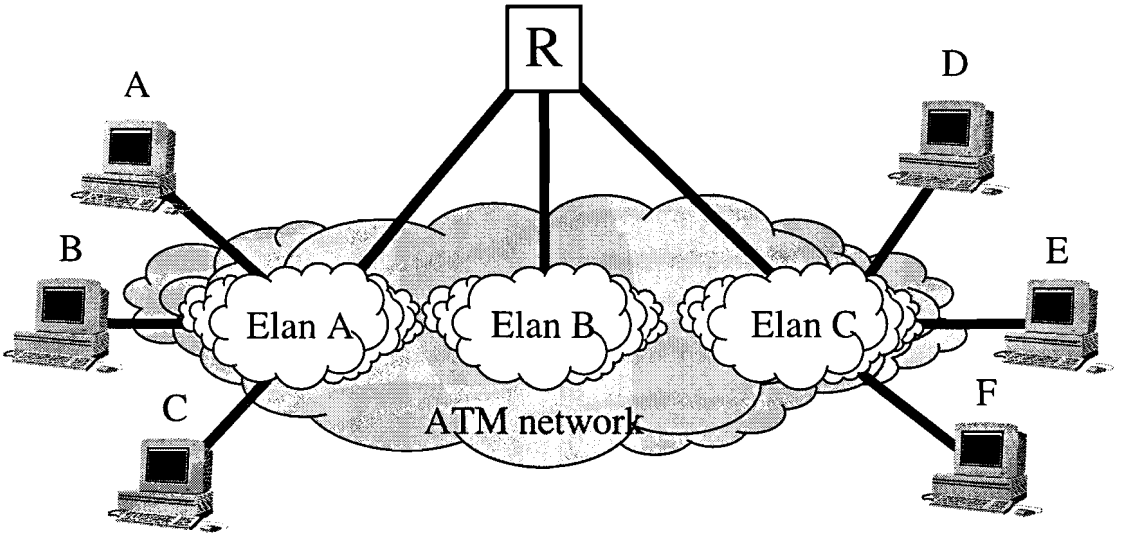


figure 6 : Situation where ATM already exists

6.3.1 NETWORKS WITH ATM SUPPORT

In Networks where ATM already is implemented in combination with Lane or Clip, Mpoa can be used as evolutionary step. figure 6 shows an example of a situation where Lane already is used. Here a lot of traffic is send through the routers and this needs to be avoided, Mpoa can be used, since it can set-up shortcuts and so by-pass routers. A Mpoa domain can be created by changing a router into a Mpoa Server. In this situation the default-path still uses Lane but the shortcut path is bypassing the router. Also there isn't much address resolution traffic, because the ingress Mpoa Server is the egress Mpoa server. Another example is given in figure 7, where a combination of Clip and Lane is used.

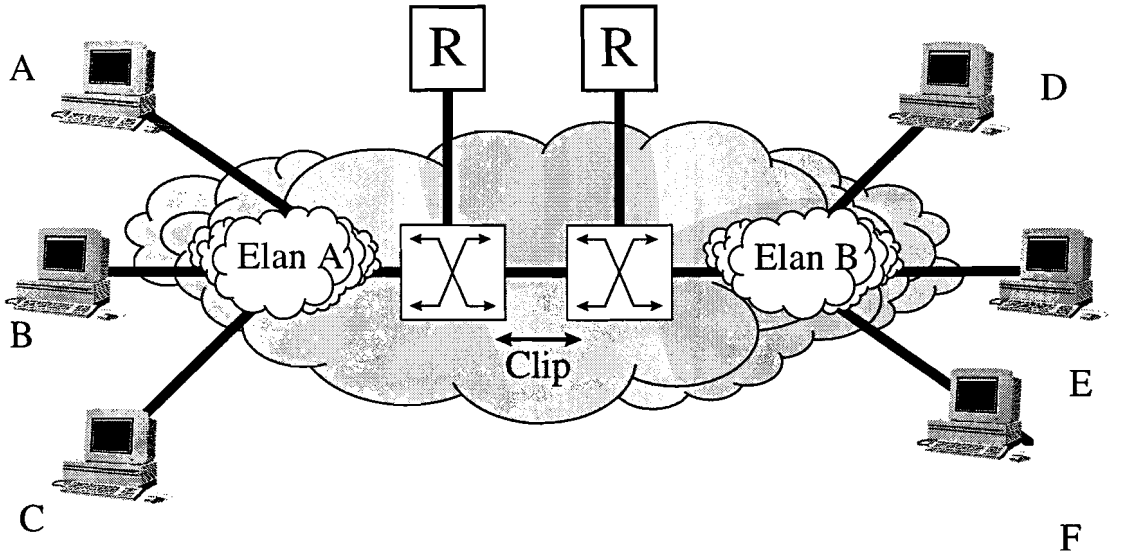


figure 7 : Situation where multiple ATM solutions already exists

6.3.2 NETWORKS WITH FREQUENT CHANGING TOPOLOGIES.

In many companies, employees don't have a static place to work. They are shifting places within the building or even within the company. An example is given in figure 8. In this figure two buildings are given. Each building consists of three levels. A employee A travels a lot within the building and has for example a portable computer, that can be connected everywhere within in the network. To make the network management easy and simple, Mpoa can be used. The virtual router solves the problem of the physical locations and makes it a virtual location.

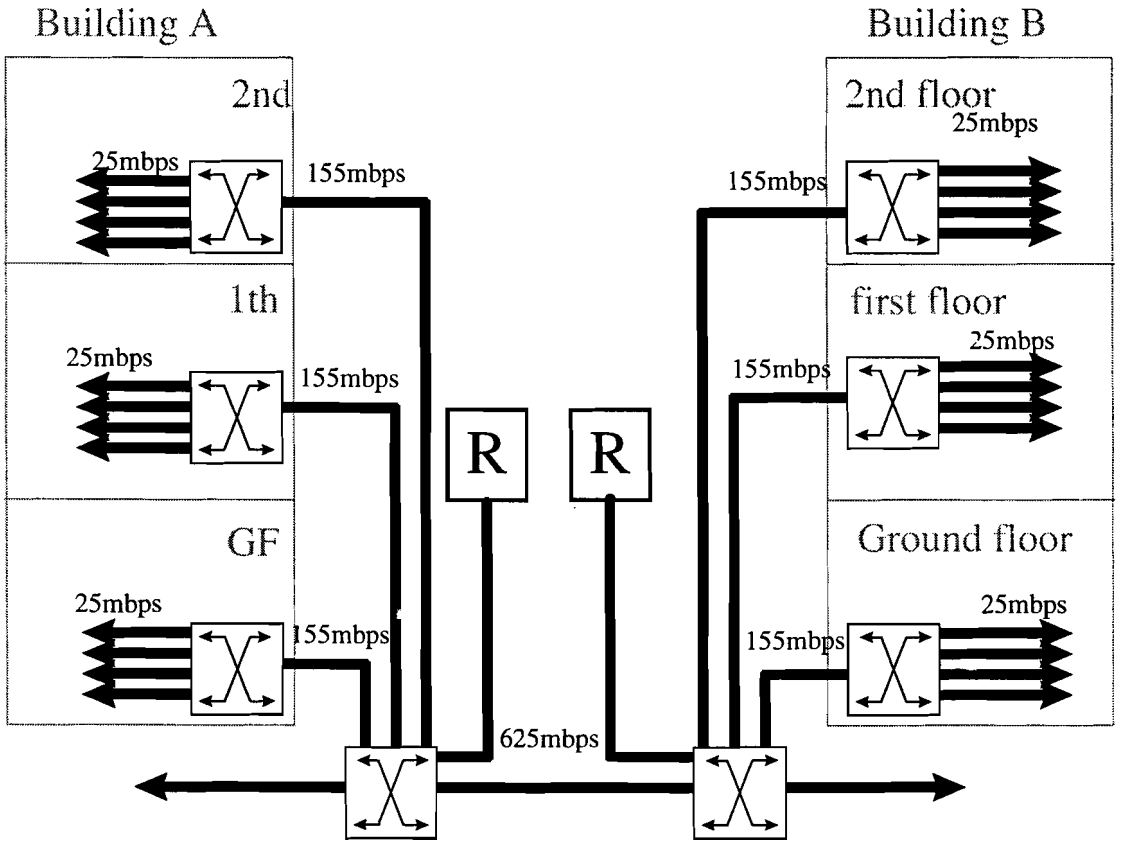


figure 8 : Example of portable work places

6.3.3 NETWORKS WHERE QOS IS NEEDED.

In networks where QoS is needed, for example for with use of picture telephone, multimedia applications, video and speech, Clip or Lane can be used. The lack of QoS of these Clip and Lane ask for Mpoa or IP switching. But when a choice must be made between IP switching or Mpoa other issues must be take into account beside QoS. These are : scaleability, capacity of the network, management and interoperability.

6.3.4 NETWORKS WITH DIFFERENT NETWORK-PROTOCOLS

In a network with different network protocols running all over the place, Mpoa can also be used. Mpoa supports IP, IPX, DECnet, AppleTalk, etc. An example of this is given in figure

9, where different networks are implemented and used in the network beside each other. It is still impossible to communicate between two clients that are using different network protocols.

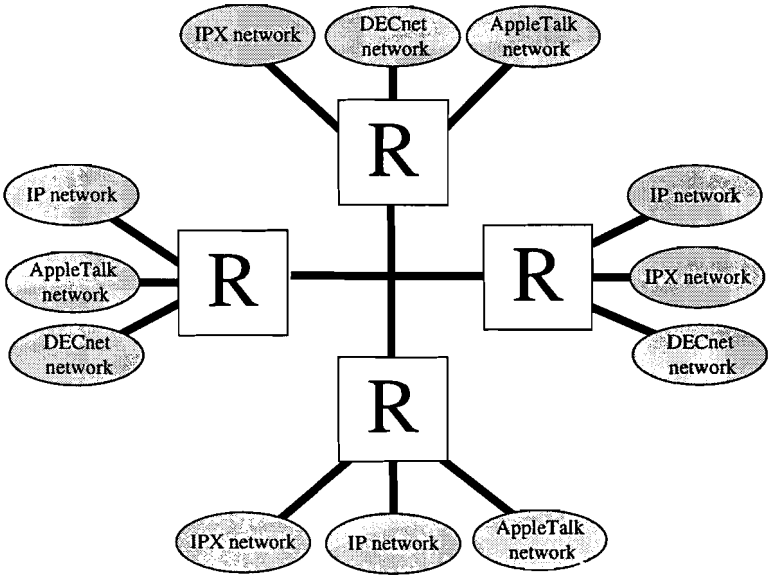


figure 9 : Network with different network protocols

6.3.5 NETWORKS WHERE ROUTER BOTTLENECK EXIST.

In situations where ATM is used, and Clip or Lane is implemented, like figure 1 to figure 4, and a lot of traffic is send through the router, the router bottleneck arises. Mpoa can be used to avoid the router for long duration flows.

Conclusion

Mpoa can be used in situations where routers need to be avoided, in situations where QoS is needed, and in situations where easy network management is needed. A disadvantage of management is the overhead in the network that arises with it.

6.4 THE USAGE OF IP SWITCHING

Mpoa and IP switching are used in the same situation. Mpoa is implemented for management reasons and IP switching for its simple use and procedures. Mpoa and IP switching both have a high performance if long duration flows are used. For short duration flows Mpoa and IP switching are ineffective. When a choice must be made between IP switching or Mpoa some issues must be take into account, such as : QoS, scaleability, capacity of the network, management, compatibilty and interoperability.

7. CONCLUSIONS & RECOMMENDATIONS

This chapter gives the answer on the questions in section 1. The questions are repeated and answered. Furthermore some recommendations are given for further research on Mpoa.

7.1 CONCLUSIONS

Research on Multi-Protocol over ATM has lead to several conclusions, as mentioned in previous chapters. In section 1, the introduction, several questions were asked by KPN regarding the research on the functionality on Mpoa. Here these questions are repeated and an answer is given.

Question 1 : "How does Mpoa work?".

Answer : Section four, deals with Mpoa. Mpoa consist of two components and several protocols to guarantee correct data traffic handling. The two components are Mpoa Servers and Mpoa Clients. Mpoa Servers take care of routing calculation in the default path and address resolution by means of an extended NHRP. There are two kinds of Mpoa Clients : an edge device and an ATM attached host. The different protocols used in Mpoa are : configuration, discovery, flow detection, target resolution, Mpoa triggers, keep-alive's, and several purges. Mpoa is a protocol that can deliver efficient communication between hosts on different networks, without passing a router. Mpoa is capable of transmitting different network protocols, i.e. IP, IPX. In the future also DECnet, Appletalk, an other protocols. Mpoa can deliver end-to-end QoS on a shortcut. Mpoa migrates bridging and routing in the same device, that is in an edge device. Finally, Mpoa makes a separation between packet forwarding and packet routing. This is called the virtual router effect.

Question 2 : "What is the relationship with other protocols or solutions?".

Answer : The other protocols and solutions are Clip, Lane, and IP switching. The relationship with Clip is, that the shortcut set-up of Mpoa looks the same as the Clip principle. This is because Clip also sets-up a direct connection. But Clip can't span route boundaries, being the difference between Clip and Mpoa. Another difference is the manageability of Mpoa. Clip doesn't support it all.
The relationship between Lane and Mpoa is very large. Lane is used as solution in the default path. You can see Mpoa as an extension of Lane, to reach clients beyond

the router boundaries. Another difference is the virtual router concept of Mpoa, which increases the manageability of Mpoa above that one of Lane.
The difference between IP switching and Mpoa is the place where flow detection takes place. Every IP switch performs flow detection, where Mpoa does the flow detection once, at the host. Another difference is the specification of the protocol. Mpoa is developed by the ATM Forum, and is an official standard when released. In contrary to IP switching that is developed by a company, and is not an official standard.

Question 3 : "How to decide when to use Mpoa and when to use other solutions?".

Answer : Other solution are used in the following situations : small networks, backbones, very static networks and in situations where no routers exists. The reason for this is the complexity of Mpoa. In small networks Clip or Lane can be used. In backbone Clip is preferred. IP switching is preferred in very static networks with routers. Lane and Clip are also preferred in intra-networks without routers. In other situations, Mpoa is the preferred solution.

Question 4 : "When is Mpoa of interest?".

Answer : Mpoa is of interest when an ATM network needs expansion beyond the router boundaries or in situations where a migration path is needed between legacy LANs and an ATM-network. In a situation where flexibility is needed, Mpoa can also be fit in. When QoS requirements are necessary for the increasing requirements of networks, like supporting video, speech, or multimedia applications, two solutions are possible : IP switching and Mpoa. The choice between these two depends on the following issues : QoS, scaleability, capacity, management, and interoperability. In networks were routers are handling a lot of traffic, Mpoa can relieve there work by bypassing them with shortcuts. The support for Mpoa in the industry is controversial. MADGE experts think Mpoa is very complicated and argue that for the support of IP-only traffic, Clip can be used. If you want to support other traffic, Mpoa can be considered. George Swallow, the chairman of the Mpoa workgroup (he's working for CISCO) argues that Mpoa will be used for flexible, local networks whereas IP-switching (and tag-switching) are more suitable for less flexible backbones.

Question 5 : "What does an implementation guideline look like?".

Answer : When Mpoa is implemented in a network with different kinds of components the following has to be changed. First the existing network has to be upgraded to an ATM network. Hereby should every device be equipped with an ATM-interface. A bridge in a 'normal' network is transferred into an edge device. Not alone the ATAM-interface should be implemented, but also the Mpoa software should be installed on the clients. This software includes a Mpoa Client and LEC. Hosts that are directly connected via an ATM-interface also need Mpoa software, including a Mpoa client and a LEC. The routers in a Network are replaced by Mpoa Servers. These Severs are equipped with a traditional router and a NHRP server and also Mpoa software.

7.2 RECOMMENDATIONS

In this thesis the main research is based on the functionality of Mpoa. The straw ballot version of Mpoa will be released in June '97, but the world of ATM is expanding fast. It is recommendable to stay in touch with the ATM Forum to monitor direction of Mpoa v2. As mentioned Mpoa is not the only protocol to transport network protocols over ATM. Therefor it is recommendable to look after other techniques as well. Beside a theoretical study on the protocols and solutions, a similar practical research has to be done, by doing a performance study on the different protocols. Hereby paying special attention to the support of short duration flows. Hereby looking at the turn-around point between default path and shortcut.

Further recommendations are :

1) TCP over ATM.

   With the techniques treaded in this document it is possible to create a smooth migration path from legacy LANs to ATM. This graduate thesis concentrates only on the use of ATM and network protocols. It is recommended to look at a migration of TCP and ATM. One of the advances of the use ATM directly under TCP instead of IP, is that of the connection-oriented nature of TCP which can be reflected directly to ATM. One of the problems with this is the use of ATM-addresses instead of IP-addresses. Most likely DNS servers must be changed. These DNS servers take care of the translation to IP-addresses.

2) Flow detection.

   Flow detection in MPOA is now based on an amount of packets passing the MPS per measuring time. This is not a completely reliable measurement because a source can influence this by varying his packet size. For example the source may send 10 packets with a size of 1500 bytes. If the source changes the packet size to 700 bytes/packet, resulting in 20 packets in the same time, it triggers the shortcut setup mechanism much faster then before, while not changing its load.

3) Bi-directional shortcuts.

   The shortcut established by Mpoa can be bi-directional. In the Mpoa specifications this I not mentioned. To support bi-directional shortcut a match should be made between ingress and egress cache. The disadvantage of this is that both cache's are no longer independent.

AppendixA : SERVICE- AND QOS-CLASSES IN ATM[1]

This table is used in section 2.5 Quality of Service, and in section 5.7 Resource reSerVation Protocol and ATMs Quality-of-Service.

QoS Class	A	B	B	C	Unspecified
ATM service Class	CBR	VBR (Real time)	VBR (non Real Time)	ABR	UBR
Connection Mode	Connection Oriented				
Timing Sensitivity	sensitive		insensitive		
Cell Loss Ratio	specified				unspecified
Cell Delay param.	specified			unspecified	
Peak Cell Rate	specified				
Burst param.	n/a	specified		n/a	
Min. Cell Rate	n/a			specified	n/a
Flow/congestion Control	no			yes	no
AAL-type	1	1?,2?,5?	3/4, 5		
Example	circuit emulation	VBR video	Connection oriented data transfer		

[1] Source R&D-RA-95-1064, "ATM-ontwikkelingen in de bedrijfsomgeving"

AppendixB : MPOA PACKET CONTENTS

1 INGRESS MPC-INITIATED MPOA RESOLUTION

Ingress MPC-Initiated MPOA Resolution includes a request phase and a reply phase. The request phase proceeds from left to right as follows:

	Ingress MPC	Ingress MPS	Egress MPS	Egress MPC
Packet Type	Mpoa Resolution Request	NHRP Resolution Request	Mpoa Cache Imposition Request	
Request ID	Request ID 1	Request ID 2	Request ID 3	
Source Protocol Address	NULL or MPC Protocol Address	I-MPS Protocol Address	E-MPS Protocol Address	
Destination Protocol Address	Destination Protocol Address	Destination Protocol Address	Destination Protocol Address	
Source NBMA Address	I-MPC Data ATM Address	I-MPC Data ATM Address	I-MPC Data ATM Address	
Client Protocol Address (1)	NULL	NULL	NULL	
Prefix Length (1)	Widest Acceptable Prefix Length	Widest Acceptable Prefix Length	Requested Prefix Length	
Holding Time				
Client NBMA Address (1)	NULL	NULL	NULL	
Extensions	Empty Mpoa Egress Cache Tag Extension Mpoa ATM Service Category Extension (1)	Received Extensions	Received Extensions Mpoa DLL Header Extension (Cache ID, ELAN ID, DLL Header)	

(1) Optional

The reply phase proceeds from right to left as follows:

	Ingress MPC	Ingress MPS	Egress MPS	Egress MPC
Packet Type		Mpoa Resolution Reply	NHRP Resolution Reply	Mpoa Cache Imposition Reply
Request ID		Request ID 1	Request ID 2	Request ID 3
Source Protocol Address		Restore to NULL[2] or I-MPC address (from original MPOA Resolution Request)	I-MPS Protocol Address	E-MPS Protocol Address
Destination Protocol Address		Destination Protocol Address	Destination Protocol Address	Destination Protocol Address
Source NBMA Address		I-MPC Data ATM Address	I-MPC Data ATM Address	I-MPC Data ATM Address
Client Protocol Address		E-MPS Protocol Address	E-MPS Protocol Address	NULL[2]
Prefix Length		Actual Prefix Length	Actual Prefix Length	Actual Prefix Length (2)
Holding Time				
Client NBMA Address		E-MPC Data ATM Address	E-MPC Data ATM Address	E-MPC Data ATM Address
Extensions		Received Extensions	Received Extensions	Received Extensions

An E-MPC can modify the Prefix Length to make it a host entry if a CIE was included in the request. An E-MPC must add a CIE with a host entry if a CIE was not included in the request.

[2] NULL: zero length, no space allocated in packet

2 EGRESS MPC-INITIATED EGRESS CACHE PURGE

Egress MPC-Initiated Egress Cache Purge includes a request phase and a reply phase. The request phase proceeds from right to left as follows:

	Ingress MPC	Ingress MPS	Egress MPS	Egress MPC
Packet Type		NHRP Purge Request	NHRP Purge Request	Mpoa Egress Cache Purge Request
Request ID		Request ID 3	Request ID 2	Request ID 1
Source Protocol Address		E-MPS Protocol Address	E-MPS Protocol Address	NULL
Destination Protocol Address		I-MPS Protocol Address	I-MPS Protocol Address	E-MPS Protocol Address
Source NBMA Address		E-MPC Data ATM Address	E-MPC Data ATM Address	E-MPC Data ATM Address
Client Protocol Address		Destination Protocol Address (to purge)	Destination Protocol Address (to purge)	Destination Protocol Address (to purge)
Prefix Length		Destination Prefix Length	Destination Prefix Length	Destination Prefix Length
Client NBMA Address		I-MPC Data ATM Address (3)	I-MPC Data ATM Address (3)	I-MPC Data ATM Address (3)
Extensions		Mpoa Egress Cache Tag Extension (3)	Mpoa Egress Cache Tag Extension (3)	Mpoa DLL Header Extension (Cache ID) (3) Mpoa Egress Cache Tag Extension (3)

The reply phase proceeds from left to right as follows:

	Ingress MPC	Ingress MPS	Egress MPS	Egress MPC
Packet Type	NHRP Purge Reply	NHRP Purge Reply	Mpoa Egress Cache Purge Reply	
Request ID	Request ID 3	Request ID 2	Request ID 1	
Source Protocol Address	E-MPS Protocol Address	E-MPS Protocol Address	NULL	
Destination Protocol Address	I-MPS Protocol Address	I-MPS Protocol Address	E-MPS Protocol Address	
Source NBMA Address	E-MPC Data ATM Address	E-MPC Data ATM Address	E-MPC Data ATM Address	
Client Protocol Address	Destination Protocol Address (to purge)	Destination Protocol Address (to purge)	Destination Protocol Address (to purge)	
Prefix Length	Destination Prefix Length	Destination Prefix Length	Destination Prefix Length	
Client NBMA Address	I-MPC Data ATM Address (3)	I-MPC Data ATM Address (3)	I-MPC Data ATM Address (3)	
Extensions	Received Extensions	Received Extensions	Received Extensions	

This field is optional. If these fields are not present in an MPOA Egress Cache Purge Request, the E-MPS has to generate an NHRP Purge Request for each I-MPS that the E-MPS has relevant Cache Entries for. Similarly , the I-MPS must generate an NHRP Purge to each affected I-MPC if these fields are not present.

3 EGRESS MPS-INITIATED EGRESS CACHE PURGE

Egress MPS-Initiated Egress Cache Purges are transacted with the Ingress MPC and the Egress MPC simultaneously. Each transaction includes a request phase and a reply phase. The request phase proceeds as follows:

	Ingress MPC	Ingress MPS	Egress MPS	Egress MPS	Egress MPC
Packet Type		NHRP Purge Request	NHRP Purge Request	Mpoa Cache Imposition Request	
Direction		←	←	→	
Request ID		Request ID 3	Request ID 2	Request ID 1	
Source Protocol Address		E-MPS Protocol Address	E-MPS Protocol Address	E-MPS Protocol Address	
Destination Protocol Address		I-MPS Addr	I-MPS Addr	Destination Protocol Address (to purge)	
Source NBMA Address		E-MPC Data ATM Address	E-MPC Data ATM Address	NULL	
Client Protocol Address		Destination Protocol Address (to purge)	Destination Protocol Address (to purge)	NULL	
Prefix Length		Destination Prefix Length	Destination Prefix Length	Destination Prefix Length	
Holding Time				0	
Client NBMA Address		I-MPC Data ATM Address (4)	I-MPC Data ATM Address (4)	NULL	
Extension				Mpoa DLL Header Extension (Cache ID) (4)	

The reply phase proceeds as follows:

	Ingress MPC	Ingress MPS	Egress MPS	Egress MPS	Egress MPC
Packet Type	NHRP Purge Reply	NHRP Purge Reply			Mpoa Cache Imposition Reply
Direction	→	→			←
Request ID	Request ID 3	Request ID 2			Request ID 1
Source Protocol Address	E-MPS Protocol Address	E-MPS Protocol Address			E-MPS Protocol Address
Destination Protocol Address	I-MPS Protocol Address	I-MPS Protocol Address			Destination Protocol Address (to purge)
Source NBMA Address	E-MPC Data ATM Address	E-MPC Data ATM Address			NULL
Client Protocol Address	Destination Protocol Address (to purge)	Destination Protocol Address (to purge)			NULL
Prefix Length	Destination Prefix Length	Destination Prefix Length			Destination Prefix Length
Client NBMA Address	I-MPC Data ATM Address	I-MPC Data ATM Address			NULL
Extensions					Received Extensions

Optional

4 DATA-PLANE PURGE

Data-Plane Purges operate in a single phase from right to left as follows:

	Ingress MPC	Egress MPC
Packet Type		NHRP Purge Request
Request ID		Unused. Set to zero
Source Protocol Address		E-MPS Protocol Address or NULL (5)
Destination Protocol Address		NULL
Source NBMA Address		E-MPC Data ATM Address
Client Protocol Address		Destination Protocol Address (to purge) (6)
Prefix Length		Destination Prefix Length
Client NBMA Address		I-MPC Data ATM Address or NULL (7)
Extensions		Mpoa Egress Cache Tag Extension (6)

(5) Use E-MPS Protocol address if purge results from an MPS dying, and NULL if purge results from an Egress Cache miss.
(6) When purging all entries associated with this E-MPS, the Client Protocol Address field is set to NULL and the Tag Extension cannot be used. (The Tag Extension and the NULL Client Protocol Address are mutually exclusive).
Optional.

5 MPOA TRIGGER

MPOA Triggers operate in a single phase from right to left as follows:

	Ingress MPC	Ingress MPS
Packet Type		Mpoa Trigger
Request ID		unused
Source Protocol Address		I-MPS Protocol Address
Destination Protocol Address		NULL
Source NBMA Address		I-MPS Control ATM Address
Client Protocol Address		Destination Protocol Address (to trigger)
Prefix Length		Destination Prefix Length
Client NBMA Address		NULL
Extensions		Mpoa DLL Header Extension (MPS Destination MAC Address only, ELAN ID)

6 MPOA KEEP-ALIVE

MPOA Keep-Alives operate in a single phase from left to right as follows:

	Egress MPS	**Egress MPC**
Packet Type	Mpoa Keep-Alive	
Request ID	Keep-Alive sequence number	
Source Protocol Address	E-MPS Protocol Address	
Destination Protocol Address	NULL	
Source NBMA Address	E-MPS Ctl Addr	
Extensions	Mpoa Keep-Alive Lifetime Extension	

AppendixC : Examples of Mpoa Control and Data Flows

1 Scenarios

A simple Mpoa network configuration is shown in Figure 1. The Mpoa network consists of two ELANs: ELAN-1 and ELAN-2. Each ELAN contains one or more Edge Devices and Mpoa Hosts. Each Edge Device supports one or more LAN hosts.

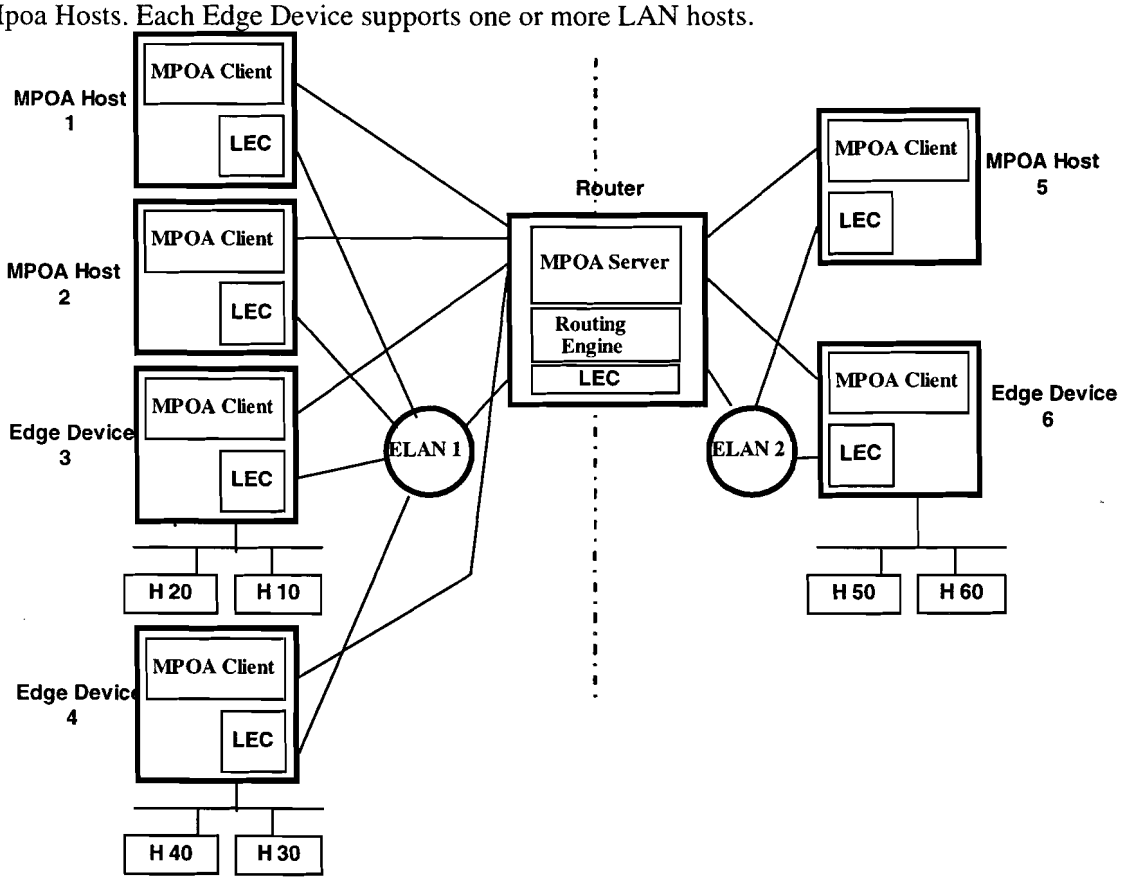


Figure 1 Example Network Configuration

To describe each flow, a source-destination pair (an MPOA Host and/or a LAN host) is chosen from Figure 1. The source and destination are chosen within the same ELAN or in different ELANs and the flows are grouped as the intra-ELAN and inter-ELAN flows.

INTRA-ELAN SCENARIOS

Intra-ELAN flows originate from an MPOA Host or a LAN host behind an Edge Device, and flow to an MPOA Host or a LAN Host in the same ELAN. These flows use LANE for address resolution and data transfer. The matrix shown in Table 1 illustrates all source-destination pairs with the matrix entry representing the scenario-index. Note that the source and destination are different hosts. The trivial scenario of a LAN-LAN flow on the same Edge Device is not covered.

Table 1 Intra-ELAN Scenarios

	To MPOA Host	**To LAN host**
From MPOA Host	(A)	(B)
From LAN host	(C)	(D)

INTER-ELAN SCENARIOS

The flows listed in Table 2 are between the source-destination pairs for which the source and destination are in different ELANs. These flows may use a default path for short-lived flows or a shortcut for long-lived flows. The default path uses the LANE and Router capabilities. The shortcut path uses LANE plus NHRP for address resolution and a shortcut for data transfer.

Table 2 Inter-ELAN Scenarios

	To MPOA Host	**To LAN Host**
From MPOA Host	(E)	(F)
From LAN Host	(G)	(H)

2 FLOWS

INTRA-ELAN

Intra-ELAN flows are illustrated in Figure 2.

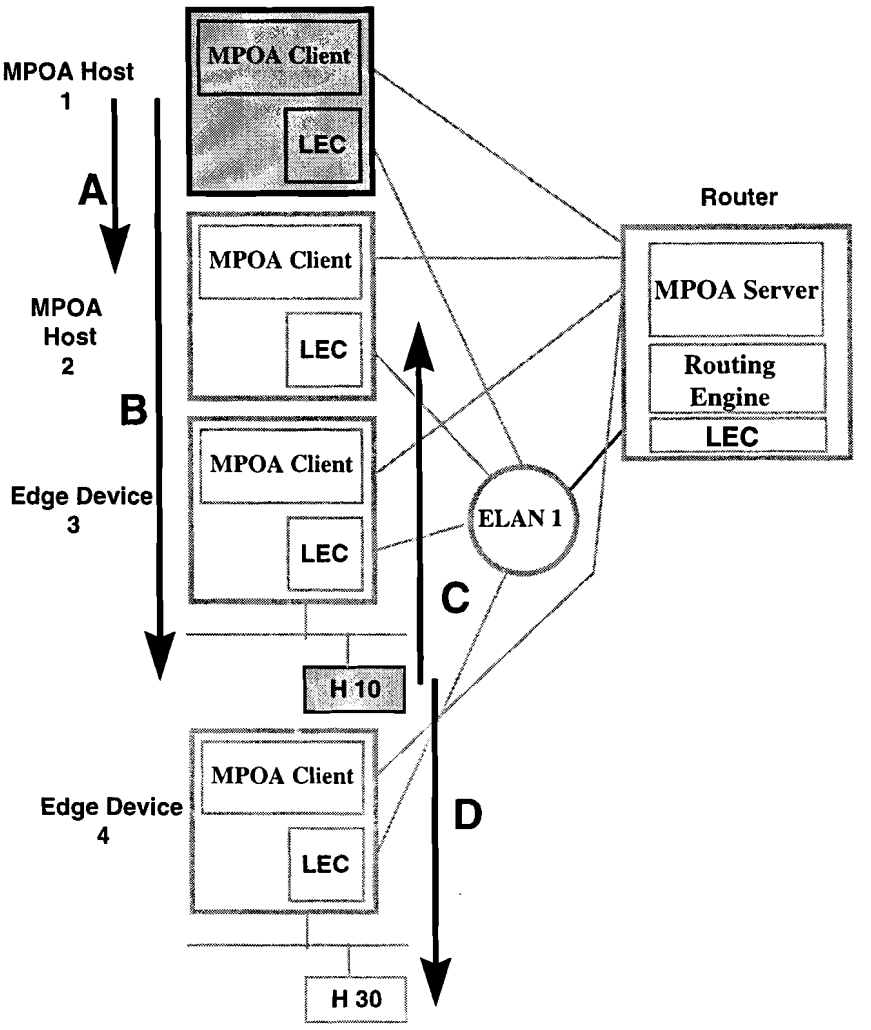


Figure 2 Intra-ELAN Flows

From MPOA Host

Scenario (A): MPOA Host 1 to MPOA Host 2

Figure 3 shows the data path for data originating from MPOA Host 1 and destined to MPOA Host 2 within the same ELAN,. LANE is used for such a flow and a Data Direct VCC will carry the LANE frames.

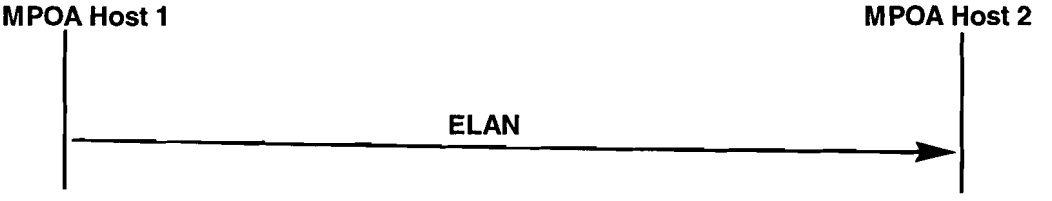


Figure 3 MPOA Host to MPOA Host Data Flow

Scenario (B): MPOA Host 1 to LAN Host H 20

Figure 4 shows the data path for data originating from MPOA Host 1 and destined to LAN Host H 20 within the same ELAN. LANE is used for such a flow and a Data Direct VCC between MPOA Host 1 and Edge Device 3 will carry the LANE frames.

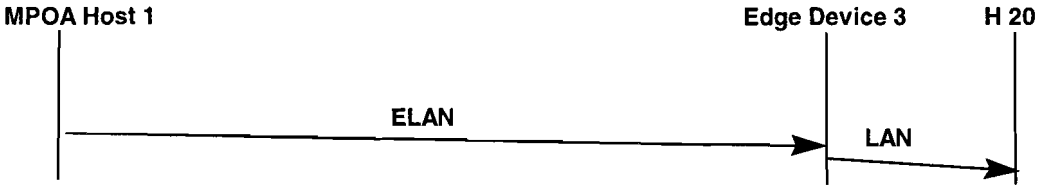


Figure 4 MPOA Host to LAN Host Data Flow

From LAN Host

Scenario (C): LAN Host H 10 to MPOA Host 2

Figure 5 shows the data path for data originating from LAN Host H 10 and destined to MPOA Host 2 within the same ELAN. LANE is used for such a flow and a Data Direct VCC between Edge Device 3 and MPOA Host 2 will carry the LANE frames.

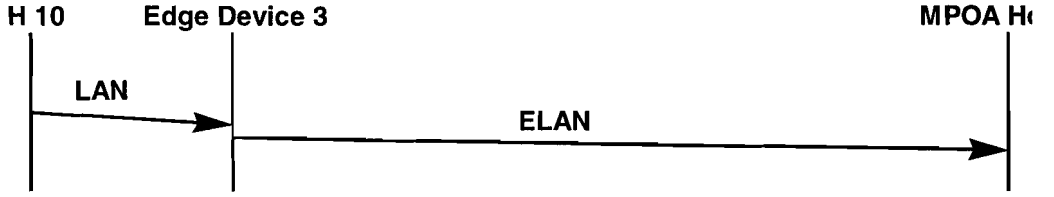


Figure 5 LAN Host to MPOA Host Data Flow

Scenario (D): LAN Host H 10 to LAN Host H 30

Figure 6 shows the data path for data originating from LAN Host H 10 and destined to LAN Host H 30 within the same ELAN. LANE is used for such a flow and a Data Direct VCC between Edge Device 3 and Edge Device 4 will carry the LANE frames.

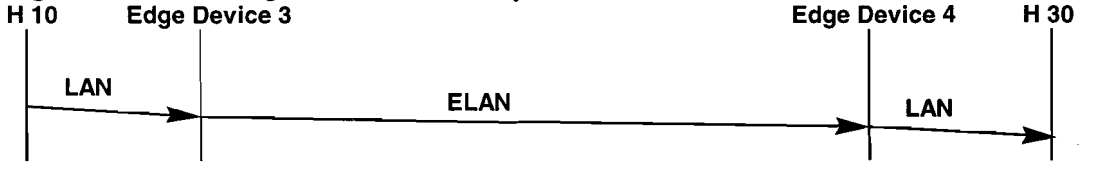


Figure 6 LAN Host to LAN Host Data Flow

Inter-ELAN

Inter-ELAN flows are illustrated in Figure 7.

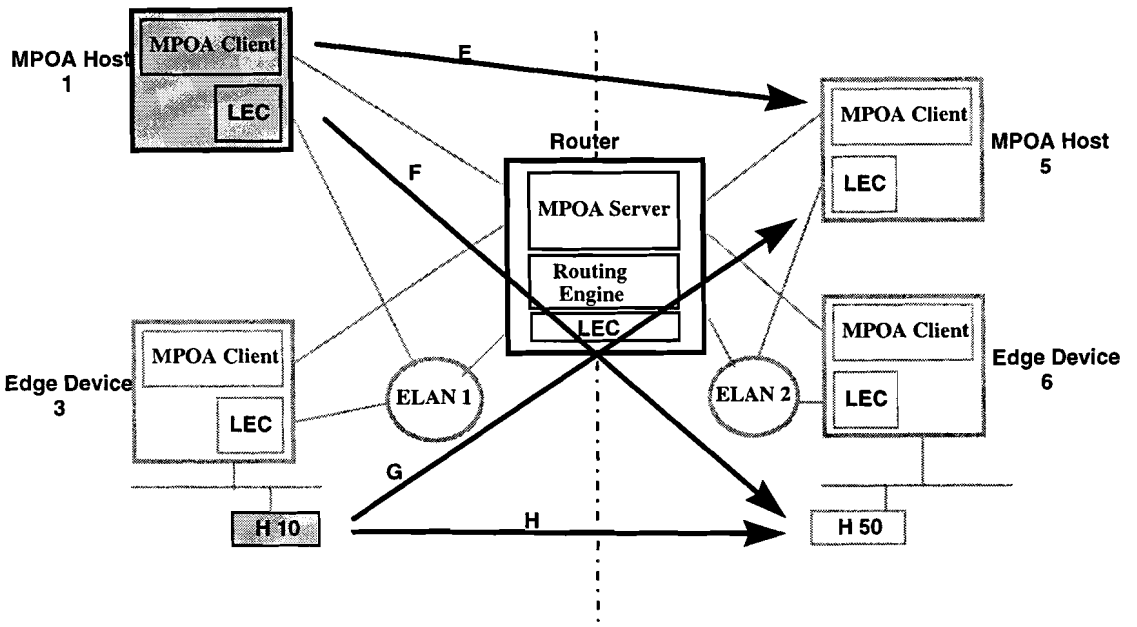


Figure 7 Inter-ELAN Flows

FROM MPOA HOST

SCENARIO (E): MPOA HOST 1 TO MPOA HOST 5

Figure 8 shows the default and shortcut data paths for data originating from MPOA Host 1 and destined to MPOA Host 5 within a different ELAN.

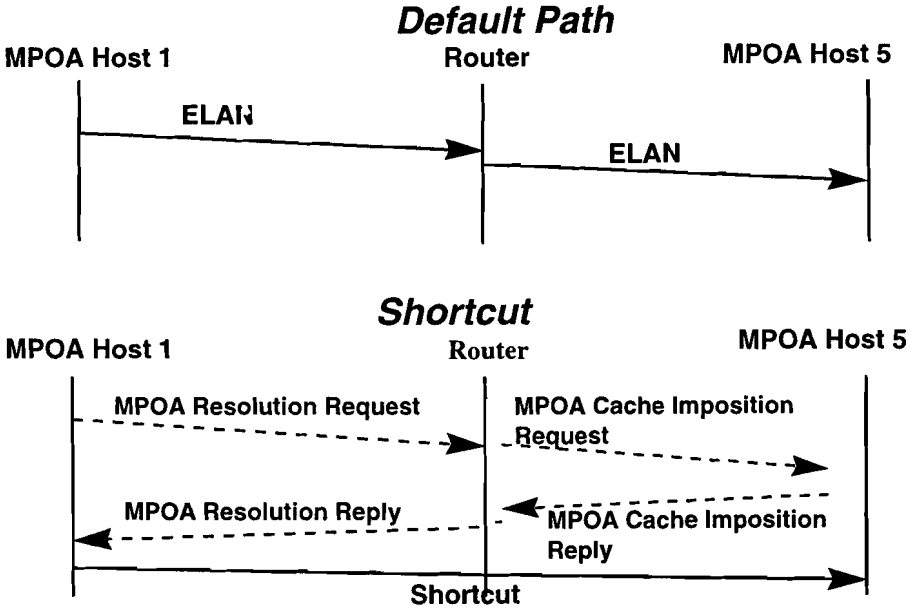


Figure 8 MPOA Host to MPOA Host

Default Path:

MPOA Host 1 sends the packet in a LANE frame to the Router via a Data Direct VCC. The Router forwards the packet in a LANE frame to MPOA Host 5 via another Data Direct VCC.

Shortcut:

If MPOA Host 1 detects a flow to the Internetwork Layer Address of MPOA Host 5, it sends an MPOA Resolution Request to the MPS to get the corresponding ATM Address. The Router sends an MPOA Cache Imposition Request to MPOA Host 5 to provide the Egress Cache Entry. MPOA Host 5 sends an MPOA Cache Imposition Reply to the MPS indicating

that it can accept the shortcut. The Router sends an MPOA Resolution Reply back to MPOA Host 1 with the ATM Address of MPOA Host 5. MPOA Host 1 may then update its Ingress Cache and establish a shortcut to MPOA Host.
For subsequent data destined to MPOA Host 5, MPOA Host 1 encapsulates the Internetwork Layer protocol packet with the appropriate encapsulation for the shortcut. The packets are then sent to MPOA Host 5 using the VCC specified in the Ingress Cache Entry.

SCENARIO (F): MPOA HOST 1 TO LAN HOST H 50

Figure 9 shows the default and shortcut data paths for data originating from MPOA Host 1 and destined to LAN Host H 50 within a different ELAN.

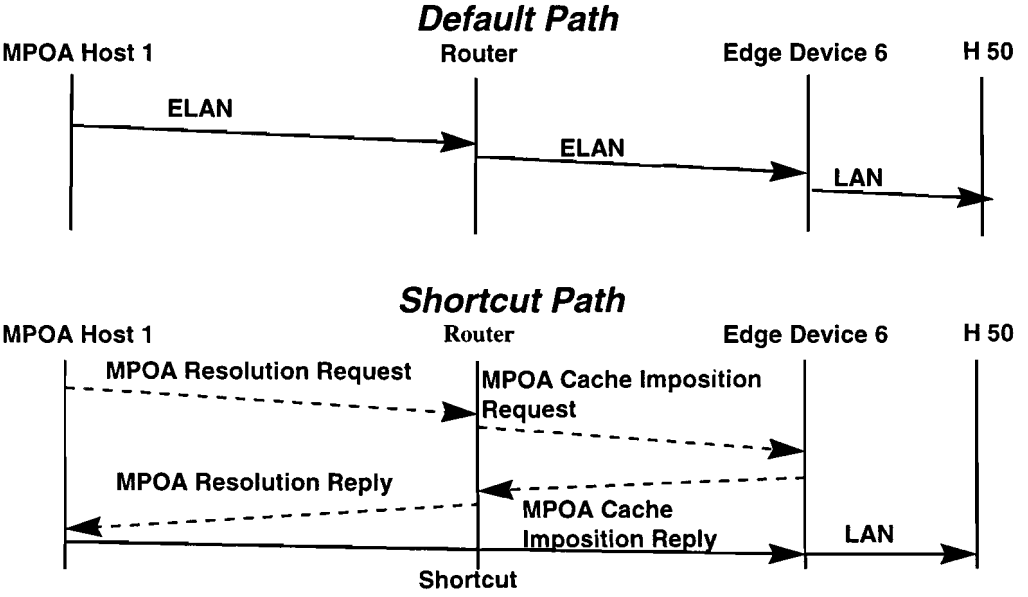


Figure 9 MPOA Host to LAN Host

Default Path:

MPOA Host 1 sends the packet in a LANE frame to the Router via a Data Direct VCC. The Router forwards the packet in a LANE frame to Edge Device 6 via another Data Direct VCC. Edge Device 6 sends the MAC frame to the LAN Host 50.

Shortcut:

If MPOA Host 1 detects a flow to the Internetwork Layer Address of LAN Host H 50, it sends an MPOA Resolution Request to the MPS to get the corresponding ATM Address. The Router sends an MPOA Cache Imposition Request to Edge Device 6 to provide the Egress Cache Entry. Edge Device 6 sends an MPOA Cache Imposition Reply to the MPS indicating that it can accept the shortcut. The Router sends an MPOA Resolution Reply to MPOA Host 1 with the ATM Address of Edge Device 6. MPOA Host 1 may then update its Ingress cache and establish a shortcut to Edge Device 6.
For subsequent data destined to LAN Host H 50, MPOA Host 1 encapsulates the Internetwork Layer protocol packet with the appropriate encapsulation for the shortcut. The packets are then sent to Edge Device 6 using the VCC specified in the Cache Entry. Edge Device 6 receives the encapsulated packets, makes the MAC frames and sends them to LAN Host 50.

FROM LAN HOST

SCENARIO (G): LAN HOST H 10 TO MPOA HOST 5

Figure 10 shows the default and shortcut data paths for data originating from LAN Host H 10 and destined to MPOA Host 5 within a different ELAN.

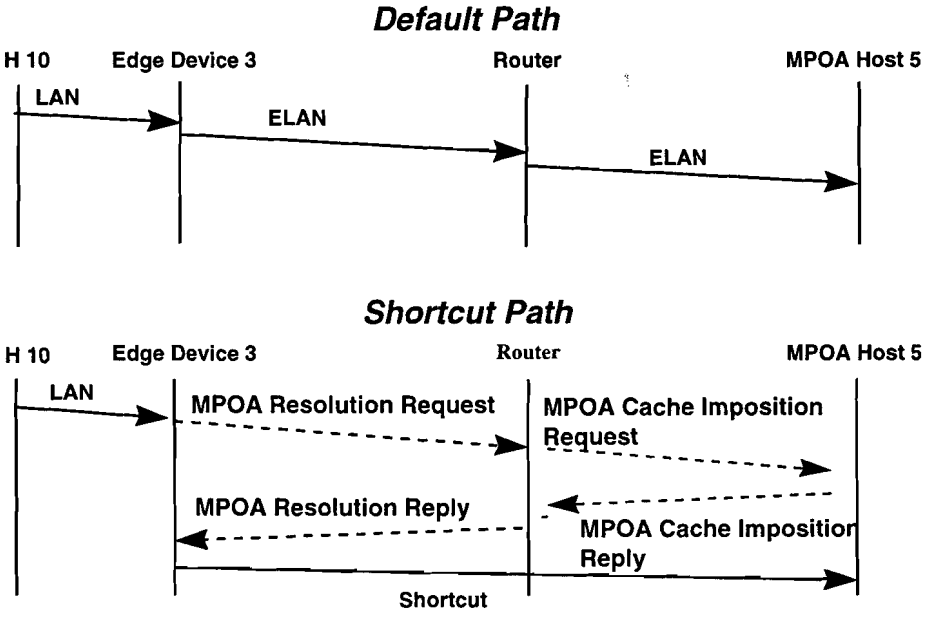


Figure 10 LAN Host to MPOA Host

Default Path:

LAN Host 10 sends the MAC frame to Edge Device 3. Edge Device 3sends the packet in a LANE frame to the Router via a Data Direct VCC. The Router forwards the packet in a LANE frame to MPOA Host 5 via another Data Direct VCC.

Shortcut.

LAN Host 10 sends the MAC frame to Edge Device 3. If Edge Device 3 detects a flow to the Internetwork Layer Address of MPOA Host 5, it sends an MPOA Resolution Request to the MPS to get the corresponding ATM Address. The Router sends an MPOA Cache Imposition Request to MPOA Host 5 to provide the Egress Cache Entry. MPOA Host 5 sends an MPOA Cache Imposition Reply to the MPS indicating that it can accept the shortcut. The Router sends an MPOA Resolution Reply to Edge Device 3 with the ATM Address of MPOA Host 5. Edge Device 3 may then update its Ingress cache and establish a shortcut to MPOA Host 5. For subsequent data destined to MPOA Host 5, Edge Device 3 encapsulates the Internetwork Layer protocol packet with the appropriate encapsulation for the shortcut. The packets are then sent to MPOA Host 5 using the VCC specified in the Cache Entry.

SCENARIO (H): LAN HOST H 10 TO LAN HOST H 50

Figure 11 shows the default and shortcut data path for data originating from LAN Host H 10 and destined to LAN Host H 50 within a different ELAN.

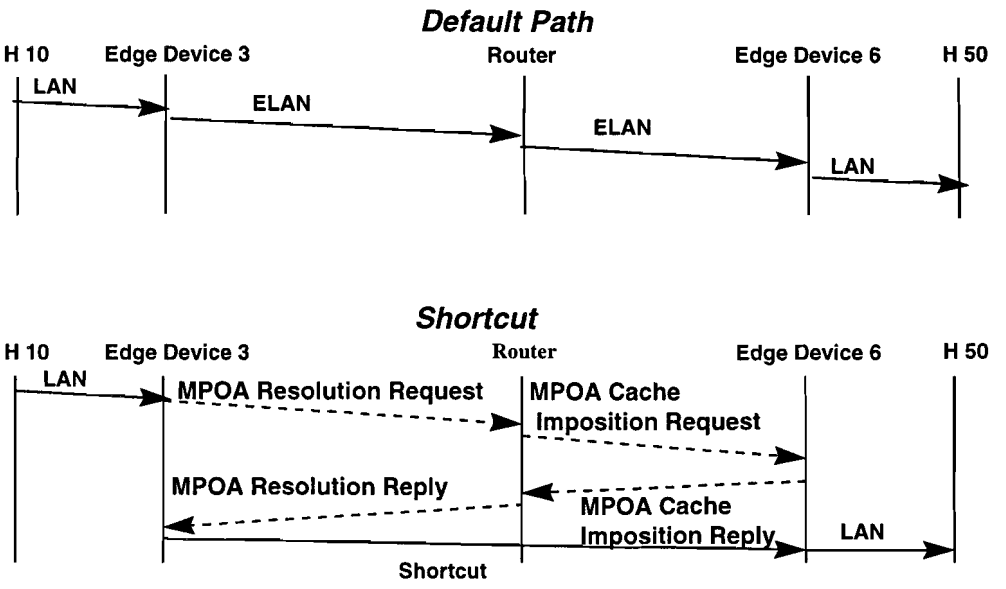


Figure 11 LAN Host to LAN Host

Default Path:

LAN Host 10 sends the MAC frame to Edge Device 3. Edge Device 3sends the packet in a LANE frame to the Router via a Data Direct VCC. The Router forwards the packet in a LANE frame to Edge Device 6 via another Data Direct VCC. Edge Device 6 sends the MAC frame to the LAN Host 50.

Shortcut:

LAN Host 10 sends the MAC frame to Edge Device 3. If Edge Device 3 detects a flow to the Internetwork Layer Address of LAN Host H 50, it sends an MPOA Resolution Request to the MPS to get the corresponding ATM Address. The Router sends an MPOA Cache Imposition Request to Edge Device 6 to provide the Egress Cache Entry. Edge Device 6 sends an MPOA Cache Imposition Reply to the MPS indicating that it can accept the shortcut. The Router sends an MPOA Resolution Reply to Edge Device 3 with the ATM Address of Edge Device 6. Edge Device 3 may then update its Ingress cache and establish a shortcut to MPOA Host 5. For subsequent data destined to LAN Host H 50, Edge Device 3 encapsulates the Internetwork Layer protocol packet with the appropriate encapsulation for the shortcut. The packets are then sent to Edge Device 6 using the VCC specified in the Cache Entry. Edge Device 6 receives the encapsulated packets, makes the MAC frames and sends them to LAN Host 50.