S-38.127 Teletekniikan Erikoistyö II **27.8.1997**
Sami Kuronen
40514H

Multicasting in IP Switching Environment

Abbreviations and Glossary

References

Abbreviations and Glossary

AAL	ATM Adaptation Layer
ATM	Asynchronous Transfer Mode
BGP	Border Gateway Protocol, Routing Protocol
B-ISDN	Broadband ISDN
BUS	Broadcast and Unknown Server, LANE Component
CLP	Cell Loss Priority, field in ATM Cell
CP	AAL Common Part
CS	AAL Convergence Sublayer
DNS	Domain Name Service
DVMRP	Distance Vector Multicast Routing Protocol
GSMP	General Switch Management Protocol, IP Switching Protocol
HEC	Header Error Control, field in ATM cell
IFMP	Ipsilon Flow Management Protocol, IP Switching Protocol
IGMP	Internet Group Management Protocol
IP	Internet Protocol
ISDN	Integrated Services Digital Network
LANE	LAN Emulation, ATM Forum Standard
LEC	LAN Emulation Client, LANE Component
LECS	LAN Emulation Configuration Server, LANE Component
LES	LAN Emulation Server, LANE Component
MPOA	Multiprotocol Protocol Over ATM
NHRP	Next Hop Routing Protocol
N-ISDN	Narrowband ISDN
NNI	Network-Node Interface, one of the two ATM Interface types
OSPF	Open Shortest Path First, Routing Protocol
PDU	Payload Data Unit
PIM	Protocol Independent Multicasting, Multicast Routing Protocol
PVC	Permanent Virtual Circuit
RIP	Routing Information Protocol, Routing Protocol
SAR	AAL Segmentation and Reassembly Sublayer
SSP	AAL Service Specific Part
SVC	Switched Virtual Circuit
TCP	Transmission Control Protocol
UDP	User Datagram Protocol
UNI	User Network Interface, one of the two ATM Interface types
VCI	Virtual Channel Identifier, field in ATM cell
VPI	Virtual Path Identifier, field in ATM cell

1 Introduction

The growth of Internet has been enormous for the past few years. The amount of multimedia traffic in the Internet has been growing all the time. Especially World Wide Web (WWW) based multimedia applications have become very popular. These facts are setting new challenges to the Internet and its protocols.

Today's multimedia applications are in many case based on unicast type distribution model that copies the same information for each recipient on the Local Area Networks (LAN) or Wide Area Networks (WAN). This wastes bandwidth and is not economical in this sense. However, there is a smarter scheme for this purpose called IP multicasting.

The new bandwidth requirements are also setting new challenges to backbone networks in WANs. Asynchronous Transfer Mode (ATM) has been proposed to be a solution to these needs. Until now the ATM standards have been driven by the ATM Forum and ITU-T. However, they have not been successful in creating an efficient solution for transporting IP traffic over ATM hardware. There has been new approaches to this purpose. They are based in standard ATM hardware, but they use proprietary solutions for the rest. One of these is IP Switching from Ipsilon.

This study will first discover the Internet protocols and networks. In the third section the principles of Ethernet networks and ATM are discussed. The fourth section deals with multimedia application types and then the fifth section discusses about multicasting in IP networks and multicast protocols. The next section introduces IP Switching and after that the study continues with multicasting in IP Switching networks.

2 Internet Protocols

In the mid 1970s, the Defense Advanced Research Projects Agency (DARPA) became interested in establishing a packet-switched network to provide communications between research institutions in the United States. With the goal of heterogeneous connectivity in mind, DARPA funded research by Stanford University and Bolt, Beranek and Newman (BBN) to create a series of communication protocols. The architecture and protocols took their current form in around 1977-1979 and the result is called the Internet Protocol suite. /1,2/

The Internet Protocols can be used to communicate across any set of interconnected networks. They are equally well suited for LANs as well as wide-area network WANs. The Internet suite includes also specifications for common applications such as mail, terminal emulation, and file transfer. Internet protocols and their relationship to the OSI reference model are presented in Figure 1. /1,2/

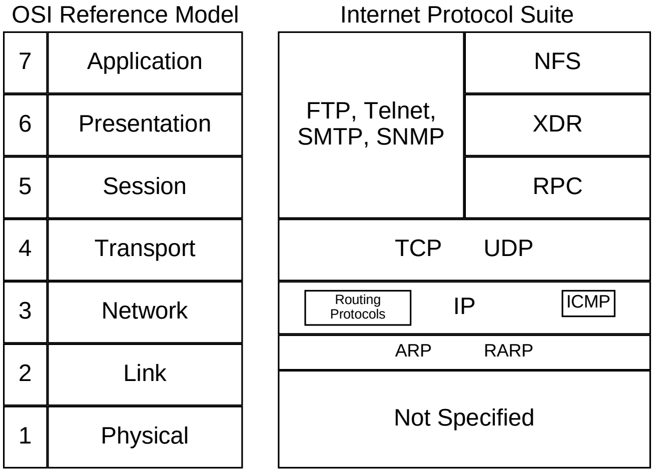


Figure 1. Internet Protocols.

2.1 IP Addressing

Each host in the Internet is assigned a unique 32-bit IP address. The IP addresses are divided logically into two parts. The first part designates the network address and the second part designates the host address. /1/

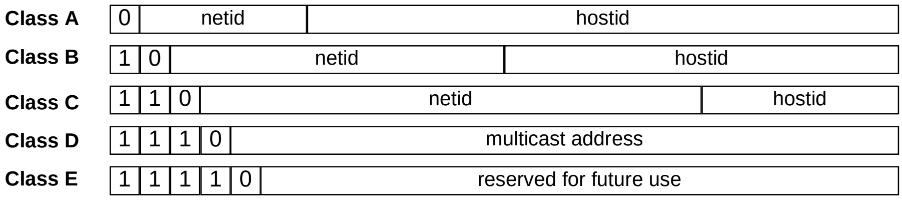

Class A 0 netid hostid
Class B 1 0 netid hostid
Class C 1 1 0 netid hostid
Class D 1 1 1 0 multicast address
Class E 1 1 1 1 0 reserved for future use

Figure 2. IP Address Classes.

IP addressing supports five different network classes, which are presented in Figure 2. A, B and C classes form the primary IP addresses, D class is for multicasting and E class is reserved for future use. /1/

When communicating with humans, the IP addresses are written as four decimal integers, which are separated by decimal points. The four integers are formed from the four octets of the IP address. This is called the dotted decimal number format of the IP address. An example of this is showed in Figure 3. In the example the IP address in dotted decimal format is 131.228.1.1, which is a class B address. /1/

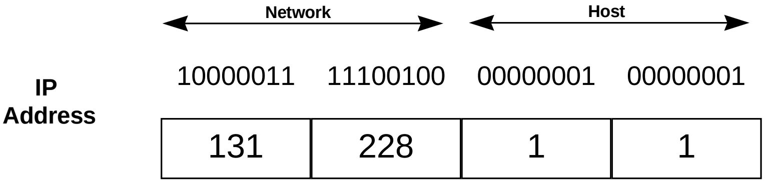


Figure 3. IP Address in Dotted Decimal Format.

Originally the IP addresses were divided into two parts; netid and hostid. However, the growth of the networks has set the need to allow a single network address to be spread over multiple physical networks. This is called subnetting. The principle of subnetting is presented in Figure 4. For example, if a company has an IP address class type B (e.g. 131.228.0.0), it can divide the address class into several class C addresses (e.g. 131.228.1.0, 131.228.2.0, …). The rest of the Internet sees only the network of class B. In other words, the decision of subnetting is local. The subnetting information is given with a subnet mask, where all the network part bits are set to one and the host bits are set to zero (e.g. in the previous example the subnet mask is 255.255.255.0). /1,2/

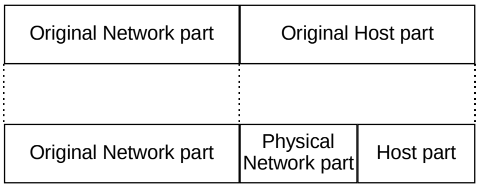


Figure 4. Subnetting.

2.2 Internet Protocol Architecture

In principle, Internet Protocols provide three sets of services. These services are shown in Figure 5. The lowest level of the stack provides connectionless delivery of packets. This forms the foundation on which everything else rests. The next level provides reliable transport service for application services. The application services form the topmost layer of Internet Protocols. This concept allows to replace one service without disturbing others. /1/

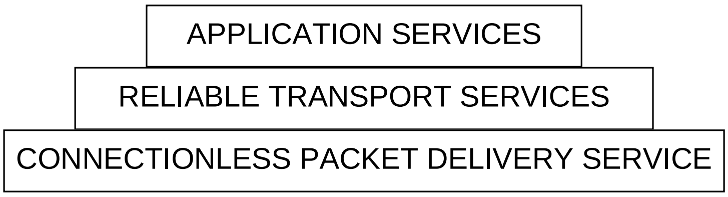

APPLICATION SERVICES
RELIABLE TRANSPORT SERVICES
CONNECTIONLESS PACKET DELIVERY SERVICE

Figure 5. Three layers of services.

2.2.1 Internet Protocol

The most fundamental Internet service consists of a packet delivery system. Technically, this service is defined as an unreliable, best-effort and connectionless packet delivery. The protocol that defines this service is called the Internet Protocol (IP). The IP protocol /1/

- defines the basic data unit to transfer data throughout the network,
- performs the routing functions and
- includes the basic rules of the unreliable packet delivery.

Basically the IP packet includes header and data, where the header gives the source and destination addresses. The IP packet is called the Internet Datagram or IP Datagram and it is showed in Figure 6. /1/

<table>
<tr><td>VERSION</td><td>HEADER LENGTH</td><td colspan="2">SERVICE TYPE</td><td colspan="4">TOTAL LENGTH</td></tr>
<tr><td colspan="4">IDENTIFICATION</td><td>FLAGS</td><td colspan="3">FRAGMENT OFFSET</td></tr>
<tr><td colspan="2">TIME TO LIVE</td><td colspan="2">PROTOCOL</td><td colspan="4">HEADER CHECKSUM</td></tr>
<tr><td colspan="8">SOURCE IP ADDRESS</td></tr>
<tr><td colspan="8">DESTINATION IP ADDRESS</td></tr>
<tr><td colspan="6">IP OPTIONS</td><td colspan="2">PADDING</td></tr>
<tr><td colspan="8">DATA</td></tr>
<tr><td colspan="8">. . .</td></tr>
</table>

Figure 6. IP Datagram.

2.2.2 Transmission Control Protocol/User Datagram Protocol

The second layer of IP Protocols provides transport services for different application services. This service can be either reliable or unreliable. The reliable transport service is called the Transmission Control Protocol (TCP), while the unreliable service is called User Datagram Protocol (UDP). The reliability of TCP is provided with a mechanism known as positive acknowledgement with retransmission. This means that the receiver sends acknowledgement messages to the sender as it receives packets. The sender keeps a record of each packet it sends and waits for acknowledgement before sending next packet. The sender starts a timer when it sends a packet and retransmits the packet if the timer expires before the acknowledgement arrives. The UDP sends packets from applications without that mechanism. /1,2/

Both TCP and UDP provide a similar way identify the upper layer applications. There are fields in the TCP and UDP packets that are called source and destination ports. The port field contains a 16-bit protocol port number, which identify the application programs at the ends of the connection. For example, TCP destination port number 23 tells that the application used is telnet (virtual terminal connection). /1,2/

2.2.3 Applications

The Internet protocol suite defines a couple of applications and protocols, such as telnet, File Transfer Protocol (FTP) and Simple Mail Transfer Protocol (SMTP). The TCP/UDP port numbers tell the application type, and currently numbers over 1023 have been assigned for well known port numbers. /1/

2.3 IP Datagram Encapsulation

As showed in Figure 1, the IP datagram is a layer 3 protocol. The standards do not define the physical network, on which IP is transported. Encapsulation refers to the manner IP datagrams carried inside network frames. /1,2/

The underlying network below IP Protocols can be based on any technology (e.g. IEEE 802.3/Ethernet, X.25, Frame Relay or ATM). The encapsulation method depends therefore on the used network type, but the principle remains the same. The principle of the encapsulation of an IP datagram inside a frame is showed in Figure 7. /1/

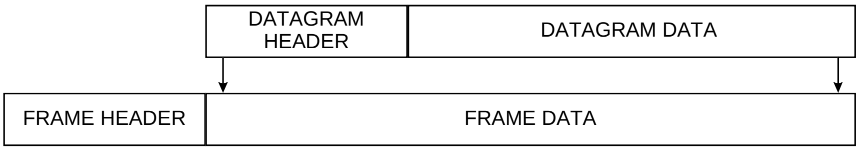


Figure 7. IP Datagram Encapsulation.

2.4 IP Routing

The Internet is composed of multiple physical networks interconnected by equipment called routers. Each router has connections to two or more networks. The hosts in contrary are typically connected only to one physical network. An example of a network is showed in Figure 8. IP routing refers to the manner IP datagrams are transported between networks. /1/

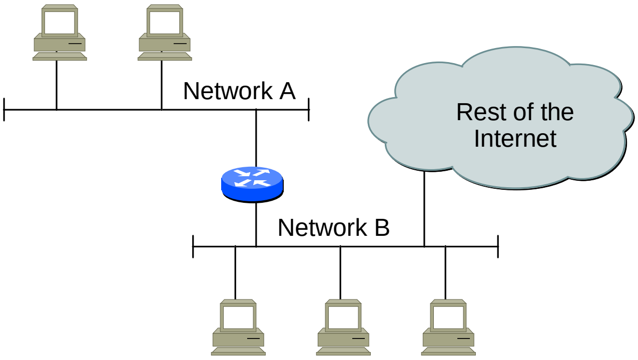


Figure 8. IP Routing

Confusion about the terms routed protocol and routing protocol is common. Routed protocols are protocols that are routed over a network, while routing protocols are protocols that route routed protocols through a network. /2/

Static routing algorithms are hardly algorithms at all. Static routes are established by the network administrator prior to the beginning of routing. They remain unchanged unless the network administrator changes them. Dynamic routing algorithms, instead, adjust to changing network circumstances in real time. /2/

Some routing algorithms operate in a flat space, while others use routing hierarchies. In a flat routing system, all routers are peers of all others. In a hierarchical routing system, some routers form a routing backbone. Packets from non-backbone routers travel to the backbone routers, where they are sent through the backbone until they reach the general area of the destination. Routing systems often designate logical groups of nodes called domains, autonomous systems or areas. In hierarchical systems, some routers in a domain can communicate with routers in other domains, while others can only communicate with routers within their domain. In very large networks, additional hierarchical levels may exist. Routers at the highest hierarchical level form the routing backbone. /2/

Some routing algorithms work only within domains, while others work within and between domains. The nature of these two algorithm types is different and therefore it is quite obvious that an optimal intradomain routing algorithm would not necessarily be an optimal interdomain routing algorithm. /2/

Link state algorithms flood routing information to all nodes in the network. However, each router sends only that portion of the routing table that describes the state of its own links. Distance vector algorithms call for each router to send all or some portion of its routing table, but only to its neighbours. In essence, link state algorithms send small updates everywhere, while distance vector algorithms send larger updates only to neighbouring routers. /2/

There are many protocols for routing IP datagrams over different IP networks. The most known ones are Routing Information Protocol (RIP), Open Shortest Path First (OSPF) and Border Gateway Protocol (BGP). RIP is a simple distance vector based routing protocol and it is also one of the oldest routing protocols. It is still used in small environments, but it not scaleable to larger networks. OSPF is a link state routing protocol that uses routing hierarchies, which are called areas. It is very common in intradomain routing. BGP is an interdomain routing protocol, that is used between domains. BGP version 4 is the most common interdomain routing protocol used at the moment. /2/

3 Network Technology Principles

3.1 Ethernet/IEEE802.3 Principles

Ethernet is the name given to the most popular packet switched LAN technology invented in the early 1970s by XEROX. Ethernet is a de facto standard released in 1978 and the equivalent IEEE version is called IEEE802.3. /1/

The original physical medium for Ethernet is a coaxial cable, but the IEEE standard defines also other physical media, such as twisted copper pair. /1,2/

Ethernet is a 10 Mbps broadcast bus technology with best effort semantics and distributed access control. All stations share a single communications channel and all attached stations receive every transmitted packet. The best effort mechanism refers to the fact that the hardware provides no information to the sender about whether the packet was delivered or not. Ethernet access control is distributed because of the access scheme called Carrier Sense Multiple Access with Collision Detection (CSMA/CD). The CSMA means that every station can access the Ethernet simultaneously and each station determines whether the Ethernet is idle by sensing the carrier wave. If two stations send simultaneously, a collision occurs. The stations can detect collisions and then try to retransmit after a while. /1/

Ethernet uses 48-bit addressing scheme. Each computer attached to Ethernet has a unique address called Ethernet address. Each manufacturer purchase blocks of Ethernet addresses and assign them to Ethernet adapters as they are manufactured. An Ethernet Address specify more than just a single destination. The address can be /1,2/

- an unicast address,
- a broadcast address or
- a multicast address.

Ethernet operates at the Data Link Layer (layer 2) of the Open System Interconnection (OSI) reference model. The data is transmitted inside frames. The Ethernet frame format is presented in Figure 9. The preamble field is a series of alternating 0s and 1s and it helps the stations to synchronise. The Frame Type field is for identifying the type of data carried in the frame. The IEEE802.3 frames differ from Ethernet frames in this field. In the IEEE802.3 frame this field is used for indicating the length of the packet. The packet type information is provided by the IEEE802.3 Logical Link Control (LLC) header followed by the SubNetwork Attacment Point (SNAP) header. IP protocols use Ethernet type frames. /1/

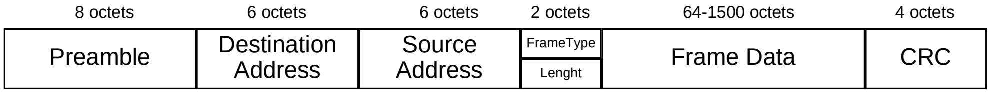


Figure 9. Ethernet/IEEE802.3 Frame.

3.2 ATM Principles

Asynchronous Transfer Mode technology is based on the efforts of the International Telecommunication Union Telecommunication Standardisation Sector (ITU-T) Study Group 13 to develop Broadband Integrated Services Digital Network (B-ISDN) for the high-speed transfer of voice, video and data through public networks. The ATM Forum has contributed a lot to the development of ATM. It was jointly founded by Cisco Systems, NET/ADAPTIVE, Northern Telecom and Sprint in 1991 to accelerate the use of ATM products and services through a rapid convergence of interoperability specifications.

3.2.1 ATM Features and Cell Structure

ATM is a packet oriented transfer mode, which uses asynchronous time division multiplexing techniques. Asynchronous in this concept refers to the manner the bandwidth is allocated among different connections. The multiplexed information flow is organised into blocks of fixed size, which are called cells. A cell consists of a 5 byte header and an 48 byte information field, where the information is carried transparently throughout the ATM network. /3,4/

ATM is also a connection-oriented technique, which means that connections are set up before any data is sent. Connection identifiers are assigned to each link of a connection when required and released when no longer needed, and they have only local significance. Signalling and user information are carried on the same physical connection, but on different virtual ATM connections. The cell header is used to identify cells belonging to the same virtual connection. ATM networks support two types of connections. Permanent Virtual Circuits (PVC) are manually configured by the network manager and Switched Virtual Circuits (SVC) are created dynamically by ATM switches using signalling. The connections can also be separated by the physical characteristics. In that sense, connections can be specified as point-to-point or point-to-multipoint. They can also be unidirectional or bi-directional. /3/

ATM connections are logically defined at two levels: virtual path (VP) and virtual channel (VC) levels. The relation between VPs and VCs is presented in Figure 10. A single physical ATM connection can include several virtual paths and virtual paths can include several virtual channels. /3,4,5/

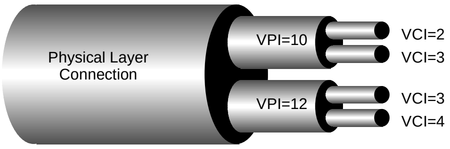
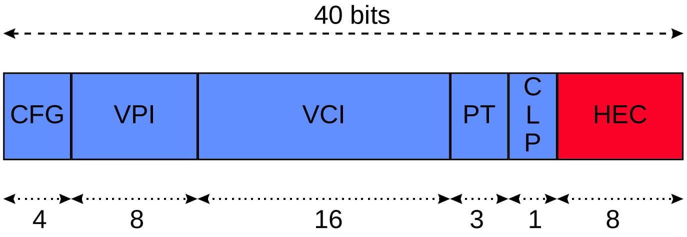


Figure 10. Virtual Paths and Virtual Channels.

The ATM standard groups have defined two header formats. User-network interface (UNI) defines communications between ATM end stations such as workstations, routers or ATM switches in private ATM networks. Network-node interface (NNI) defines communications between ATM switches. The format of the UNI cell header is shown in Figure 11. /3,4,5/



GFC	Generic Flow Control
VPI	Virtual Path Identifier
VCI	Virtual Channel Identifier
PT	Payload Type
CLP	Cell Loss Priority
HEC	Header Error Control

Figure 11. UNI Cell Header.

The first field Generic Flow Control (GFC) is purposed for flow control. This field is very rarely used at the moment and the use of it has even not been specified by the standardisation bodies. This field is missing from the NNI cells. /3,5/

The Virtual Path Identifier (VPI) and Virtual Channel Identified (VCI) fields constitute a label to identify the virtual connection at the virtual path and channel levels. Certain VPI/VCI values are reserved for special uses e.g. signalling, OAM flows, ILMI. The VPI and VCI values have only local significance and they are translated at each ATM switch. /3,4,5/

The PT field is mainly purposed for separating cells containing user data and network information. For user data cells, the field can also be used by the network to indicate that congestion has been detected. /3,5/

The CLP bit is meant for two level loss priority for individual cells (CLP=0 for high and CLP=1 for low). /3/

The HEC field uses a cyclic redundancy check for error protection of the cell header. The information field is not protected against errors at this level. The HEC field can correct single errors and detect multiple. Cells with detected but uncorrected header errors are discarded. /3,5/

3.2.2 The Layered Model of ATM

The ITU-T has defined a layered model of ATM, which is called the B-ISDN protocol reference model (PRM). It defines the functions associated with individual layers of B-ISDN. The PRM is shown in Figure 12. /6/

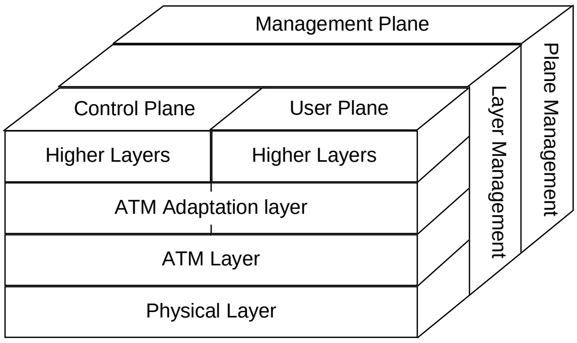


Figure 12. B-ISDN Protocol Reference Model/I.321.

In the B-ISDN reference model, the ATM Layer and the ATM Adaptation Layer (AAL) are roughly analogous parts of the Data Link Layer (layer 2) of the seven layered OSI reference model. The Physical Layer of the B-ISDN reference model is also analogous to the Physical Layer (layer 1) of the OSI reference model. /5,6/

Physical Layer

The ATM physical layer controls transmission and receipt of bits on the physical medium. It also keeps track of ATM cell boundaries and packages cells into the appropriate type of frame for the physical medium being used. /5,7/

ATM Layer

The ATM layer is responsible for establishing connections and passing cells through the ATM network. It uses the information contained in the header of each ATM cell. The data inside the payload fields of each cell is carried transparently through the ATM network. /5,7/

ATM Adaptation Layer

ATM cells are supposed to offer a simple way to carry data and therefore there has to be a way to support data transport for other more complicated sources (e.g. continuous bit rate, video, connectionless data). The ATM adaptation layer (AAL) translates between the larger service data units (SDUs) of upper-layer processes and ATM cells. The ATM adaptation layer receives packets from upper-level protocols (such as IP protocols) and breaks them into the 48-byte segments that form the payload field of an ATM cell. /5,7/

The AALs are generally divided into two parts: Segmentation and Reassembly (SAR) and Convergence Sublayer (CS). /5,7,8/

ITU-T has specified four ATM adaptation layers, but one of these is the most commonly used. The AAL5 is a simple adaptation layer that supports variable bit-rate services with error detection above the ATM layer. /5,8/

3.2.3 ATM AAL5 and IP Datagram Encapsulation

AAL5

AAL5 consist of a common part (AAL5 CP) and a service-specific part (AAL5 SSP). /5/

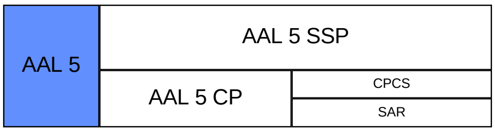


Figure 13. ATM Adaptation Layer 5.

The AAL5 CP provides unguaranteed connection-oriented transport of variable-length data packets with error detection. The common part of AAL5 consists actually of convergence sublayer and segmentation and reassembly sublayer. The SAR splits the AAL5 CPCS PDU into 48 byte segments without adding any additional information. These units are then carried in the information fields of an ATM cell stream. The reverse AAL5 functions are performed at the receiver side to reconstruct the CPCS PDU. The user data frames can be up to 65 535 bytes in length. The CP PDU is shown in Figure 14. /5,7/

The AAL5 SSP provides additional functions as required by the higher layers. In some cases it is not needed at all. An example of service-specific part is the Signalling ATM Adaptation Layer (SAAL) service specific convergence sublayer, which provides the way to transfer signalling information inside ATM cells. /5,7/

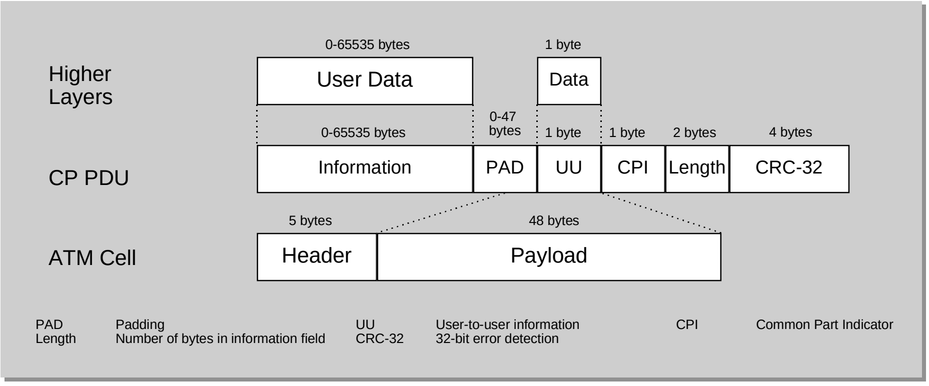


Figure 14. AAL5 CP PDU.

The CRC and length fields are used to detect errors or information loss. The detected errors are then reported to higher layers, which are responsible for the recovery. The UU and CPI fields are currently unused. The information carried inside the CP PDU can be such in length that the segmented CP PDU would not be an exact multiple of 48 octets. The PAD field is used for this purpose. /5/

LLC/SNAP Encapsulation

When IP packets are sent over Ethernet networks, the Ethernet frame contains a field called Frame Type to identify the protocol carried inside the frame (see Figure 9). The AAL5 does not provide this kind of field. There are then two possibilities: /9/

- certain VC is used to carry predefined protocols
- certain bytes inside the information field of AAL5 CP PDU is used to identify the carried protocol

The standards say that it is possible to use both methods. When the latter method is used, the protocol type information is provided by the standard IEEE802.3 LLC header followed by the SNAP header. This encapsulation method is called LLC/SNAP encapsulation. The details can be seen in Figure 15 and Figure 16. The LLC value AA.AA.03 indicates that there is a SNAP header to follow and the SNAP type value 08.00 indicates that the carried protocol is IP. /1,9/

<table>
<tr><th colspan="2">0</th><th>16</th><th>31</th></tr>
<tr><td colspan="3">LLC (AA.AA.03)</td><td>OUI$_1$ (00)</td></tr>
<tr><td colspan="2">OUI$_2$ (00.00)</td><td colspan="2">TYPE (08.00)</td></tr>
<tr><td colspan="4">IP DATAGRAM</td></tr>
<tr><td colspan="4">...</td></tr>
</table>

Figure 15. LLC/SNAP Header and IP Datagram.

The IP packet preceding the LLC/SNAP header is then carried inside the information field of AAL5 CP PDU like showed in Figure 16. /1,9/

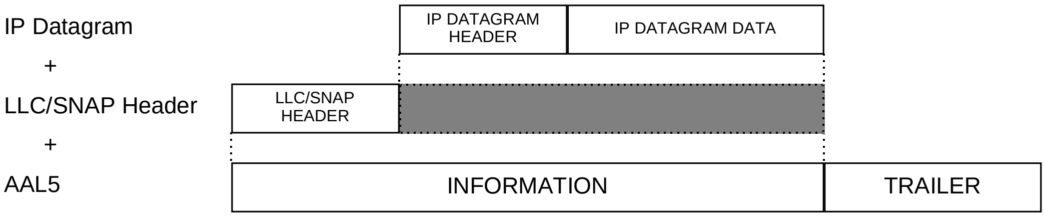


Figure 16. LLC/SNAP Encapsulation over AAL5.

3.2.4 LANs and ATM

Due to the difference between LANs and ATM, there must be a way to support coexistence between these networks. The most common network layer protocol used at the moment is IP. There are currently many alternatives for transporting IP traffic over ATM networks. /1,2,5,9,10/

Classical IP over ATM emulates IP subnets over ATM hardware by creating Logical IP subnets. This allows the IP subnet to be physically separated over the ATM network. /1/

ATM Forum LAN Emulation v. 1.0 (LANE) emulates connectionless Ethernet/TokenRing LANs over ATM. These emulated LANs form virtual LANs over ATM clouds. The basic components of LANE are LAN Emulation Server (LES), LAN Emulation Configuration Server (LECS), Broadcast and Unknown Server (BUS) and LAN Emulation Client (LEC). The first three components form the LAN Emulation Service, which the LEC uses. The LEC is one part of Virtual LAN, which can be thought as an Ethernet segment, for example. If the LEC wants to communicate between other LECs that are part of other Virtual LAN, the LEC needs to use traditional router that is connected to both Virtual LANs. /9/

Multiprotocol over ATM (MPOA) is a more sophisticated solution from ATM Forum. It uses LAN Emulation inside IP subnets, but it uses Next Hop Routing Protocol (NHRP) between IP subnets. NHRP is a short-cut routing scheme, which identifies traffic flows and then creates a short-cut between the end stations, and in the ATM world this means SVC. /10/

The most recent approaches to these needs are non-standard solutions based on the intelligence of IP protocols and to the performance of ATM switching. There are two dominating technologies at this segment. The first one is IP Switching from Ipsilon Networks. The second solution of this category is Tag Switching from the most dominant player of the routing market, Cisco Systems. This study will handle the IP switching more thoroughly in the Section 6. /10/

4 Multimedia Application Types

Nearly all desktop and also laptop computers are nowadays multimedia capable. The high performance and the new capabilities of PCs have spawned a new class of multimedia applications. The growth of Internet and its most popular application World Wide Web (WWW) has put multimedia applications into networking infrastructure to deliver live video and audio applications to end users. /12/

Basically there are three types of multimedia applications: unicast, broadcast and multicast. /12/

4.1 Unicast Type Applications

Unicast type applications send a copy of each packet to every host that want to receive the packet. This kind of application is quite easy to implement, but it has some disadvantages. First of all, this type of application requires extra bandwidth from the network, because the network has to carry the same packet multiple times. The number of receivers is also limited to the number of copies of packets that can be made by the CPU that runs the unicast application. Many applications nowadays are still unicast type. /2,12/

4.2 Broadcast Type Applications

Broadcast applications send each packet to a broadcast address. This type of application is even simpler to implement than a unicast application, but it can have serious effects on the network. If the broadcasts are allowed to travel throughout the network, the network can be heavily loaded with broadcast traffic although only a limited number of users use that service. The hosts in the network can also be heavily loaded, because they have a lot of packets to process (will the packet be discarded or not). /2,12/

4.3 Multicast Type Applications

Multicast applications send each packet to a multicast group address. Hosts that want to receive that packets indicate that they want to be members of the multicast group (i.e. they want to become leaves of the multicast tree). Multicast applications and underlying multicast protocols control multimedia traffic and shield hosts from having to process unnecessary broadcast traffic. This kind of multimedia applications are obviously the most interesting ones. /2,12/

5 Multicasting in IP Networks

The Internet Protocol suite was originally designed for communications between two computers using unicast addresses. To send a message to all devices connected to the network, a single network device uses a broadcast address. These two forms of addressing have been sufficient for transferring traditional data (such as files and virtual terminal connections). /1/

Application developers are trying to deliver data (such as the live audio and video streams) to group of devices connected to the network and another form of addressing is required. The new form of addressing is called multicast addresses, and it involves the transmission of a single IP datagram to multiple hosts. Besides addressing, IP networks require additional multicast protocols for interchanging multicast routing information and also applications need to have a way to join a multicast group. /1,2/

The Internet Engineering Task Force (IETF) has developed standards that address the required needs to support multicast communications in IP networks. The support comes with: /1,2,12/

- Multicast IP addressing
- Dynamic registration to multicast groups
- Multicast routing protocols

5.1 Multicast IP Addresses

IP address classes were discovered in Section 2.1. IP multicasting applications use class D addresses to address packets. Class D address format is showed in Figure 17.

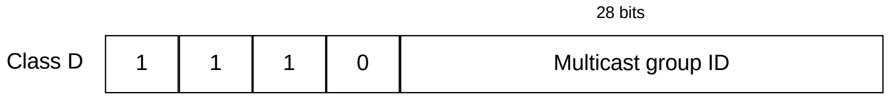


Figure 17. Multicast IP Address Class.

The class D IP addresses differ from A, B and C class addresses in the way that the last 28 bits of the address have no structure. The multicast group address is the combination of the high order bits of 1110 and the multicast group ID. Written in dotted decimal format, multicast addresses are in the range 224.0.0.0-239.255.255.255. /1/

The set of hosts that respond to a particular IP multicast address is called a host group. The group can span multiple networks. Membership in a host group is dynamic, so hosts can join and leave host groups. /1,2,12/

5.2 Multicasting in Underlying Hardware

5.2.1 Multicasting in Ethernet Networks

Ethernet packets use 48-bit (6 bytes) addressing as described in Section 3.1. They are often written in hexadecimal (e.g. 01:23:45:67:AB:CD). Addresses are usually contained in the hardware on the interface cards. The first 3 bytes (24 bits) of the addresses are specified by the IEEE on a vendor-dependent basis, while the last 3 bytes (24 bits) are specified by the Ethernet or IEEE 802.3 vendor. The source address is always a unicast address, while the destination address may be unicast, multicast, or broadcast. The first byte of an Ethernet address defines whether the address is unicast or multicast. In multicast addresses, this byte is set to 01. In the broadcast address all the bits in the address are set to 1, i.e. in hex FF:FF:FF:FF:FF:FF. /1,12/

The Internet Assigned Numbers Authority (IANA) owns a block of Ethernet addresses; in hexadecimal 00:00:5E. It includes addresses in the range 00:00:5E:00:00:00 - 00:00:5E:FF:FF:FF. The IANA has allocated half of this block for IP multicasting in the range of 01:00:5E:00:00:00 - 01:00:5E:7F:FF:FF. The first 25 bits of the address form a bit sequence 00000001 00000000 01011110 0. The last 23 bits in the Ethernet address correspond to the IP multicast group ID. The mapping is showed in Figure 18. The upper 5 bits in the IP multicast group ID are ignored in this mapping, so 32 multicast group IDs map to one Ethernet multicast address. The receiving host must therefore perform filtering functions for multicast frames. /1,12/

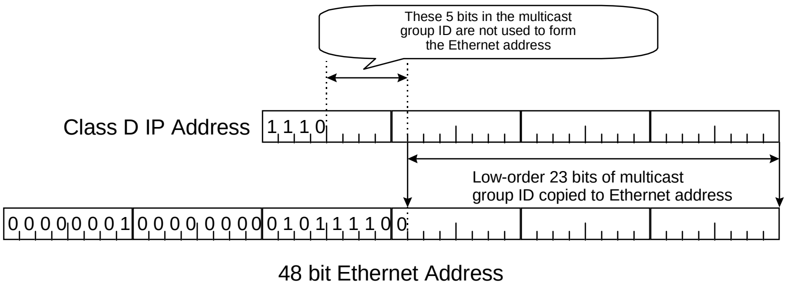

These 5 bits in the multicast group ID are not used to form the Ethernet address
Class D IP Address
1 1 1 0
Low-order 23 bits of multicast group ID copied to Ethernet address
0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 1 0 1 1 1 1 0 0
48 bit Ethernet Address

Figure 18. Multicast Address Mappings.

5.2.2 Multicasting in ATM Networks

ATM supports multicasting with point-to-multipoint connections as described earlier. However, when thinking of running LAN protocols (such as IP) over ATM, it would be desirable for ATM to support multipoint-to-multipoint links, which would be equivalent to a broadcast VCC. However, the AAL5 does not provide a way for the receiver to identify individual cells from specific sources when cells are interleaved from multiple sources. This would not allow proper reassembly of cells into frames, unless all of the cells of a specific frame are sent in proper order. /9/

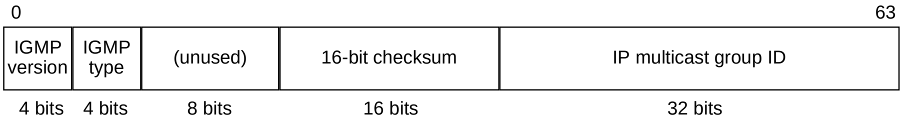
A multicast server is one solution to this problem. A multicast server can exist in an ATM network, and all members of a multicast group can establish point-to-point VCs to it. The multicast server would then create a point-to-multipoint VCC to all members on the group, with itself at the root. The receivers act as leaves. Any data sent to the multicast server is serialised, sent out the point-to-multipoint tree and received by the members of the group. The value of this approach is that cells from different sources are serialised and sent in order rather than interleaved. The multicast server can also support dynamic groups because members can be added and deleted as leaves on the tree. /9/

The multicast server approach to support multicasting in ATM networks is used in LAN Emulation, for example. The disadvantages of this approach are the need for a separate equipment, poor performance and a single point of failure. /9/

5.3 Internet Group Management Protocol

The process of joining dynamically a multicast group is fundamental to multicasting. The Internet Group Management Protocol (IGMP) is used for IP multicast group registration. /1,12/

The IGMP is part of the IP layer and it uses IP datagrams to transmit information about multicast groups. The IGMP message consists of a 20-byte IP header and 8-byte IGMP message. IGMP messages are specified in the IP datagram with a protocol value of 2. The IP packet format is showed in Figure 6 and the IGMP message format in Figure 19. /1,12/



0				63
IGMP version	IGMP type	(unused)	16-bit checksum	IP multicast group ID
4 bits	4 bits	8 bits	16 bits	32 bits

Figure 19. IGMP Message Format.

5.4 Multicast Routing Protocols

Routers need to change routing information in order to route the IP datagrams with multicast addresses to appropriate end stations.

5.4.1 Distance Vector Multicast Routing Protocol

Distance Vector Multicast Routing Protocol (DVMRP) uses a technique called reverse path flooding. With reverse path flooding, on receipt of a packet, the router floods the packet out all paths except the path that leads back to the source of the packet, which insures that a data stream reaches all LANs. If the router is attached to a LAN that does not want to receive a particular multicast group, the router sends a "prune" message back to the source to stop the data stream. When running DVMRP, routers periodically reflood the network to reach new hosts, using an algorithm that takes into account the frequency of flooding and the time required for a new multicast group member to receive the data stream. /1,2,12/

To determine which interface leads back to the source of a data stream, DVMRP implements its own unicast routing protocol. The DVMRP unicast routing protocol is similar to RIP. The path that multicast traffic follows may not be the same as the path that unicast traffic follows. /2,12/

The need to reflood prevents DVMRP from scaling well. In spite of its limitations, DVMRP is widely deployed in the IP research community. It has been used to build the multicast backbone (MBONE) across the Internet. /2,12/

5.4.2 Multicast Open Shortest Path First

Multicast Open Shortest Path First (MOSPF) is an extension to OSPF. OSPF is a unicast routing protocol that requires each router in a network to be aware of all available links in the network. Each OSPF router calculates routes from itself to all possible destinations. MOSPF works by including multicast information in OSPF link states. MOSPF calculates the routes for each source/multicast group pair when the router receives traffic for that pair. These routes are cached until a topology change occurs, which requires MOSPF to recalculate the topology. /1,2,12/

MOSPF works only in internetworks that are using OSPF and is best suited for environments in which relatively few source/group pairs are active at any one time. MOSPF performance degrades in environments that have many active source/group pairs and in environments in which links are unstable. /2,12/

5.4.3 Protocol Independent Multicast

Multicast traffic tends to fall into one of two categories: traffic that is intended for almost all LANs and traffic that is intended for relatively few LANs. Protocol Independent Multicast (PIM) is an Internet draft by Cisco Systems that has two modes of behaviour for the two traffic types: dense mode and sparse mode. A router that is running PIM can use dense mode for some multicast groups and sparse mode for other multicast groups. /2,12/

Dense Mode

In dense mode, PIM uses reverse path flooding and is similar to DVMRP. One significant difference between PIM and DVMRP is that PIM does not require a particular unicast protocol to determine which interface leads back to the source of a data stream. Instead, PIM uses whatever unicast protocol the network is using. /2,12/

Sparse Mode

In sparse mode, PIM is optimised for environments in which there are many data streams but each data stream goes to a relatively small number of the LANs in the network. For this type of traffic, reverse path flooding wastes bandwidth. /2,12/

PIM-SM works by defining a rendezvous point. When a sender wants to send data, it first sends to the rendezvous point. When a host wants to receive data, it registers with the rendezvous point. Once the data stream begins to flow from the sender, to the rendezvous point, and to the receiver, the routers in the path optimise the path automatically to remove any unnecessary hops, including the rendezvous point. /2,12/

6 IP Switching

IP traffic in the Internet and also in private networks has been growing exponentially for some time. This growth is beginning to stress the traditional routers, whose power have been processor based. Switching technology offers much higher bandwidth, but it has been available only for bridging LAN protocols (such as Ethernet). This means switching in the data link layer of the OSI reference model. Various solutions have been proposed for supporting routing over ATM switching technology. However, they have not been able to provide an efficient solution for IP routing.

Ipsilon Networks Inc. has proposed an alternative solution to combine the performance of ATM and the intelligence of IP protocols, which is called IP Switching. The aim of it is to combine the flexibility of IP with the speed of switching at low cost. /13/

The principles of IP Switching are presented in this section.

6.1 The Concept of IP Switching

The section 2 defined the IP as connectionless protocol and the section 3.2 defined ATM as a connection oriented network technology. IP switching tries to combine these two technologies together without sacrificing the scalability and flexibility. The other solutions of implementing IP over ATM obscure the real topology of the underlying network from the network layer routing protocol, which means that the underlying ATM network becomes a opaque cloud for IP. There is also a duplication of functionality, for example in routing. Multicasting capability in these solutions is quite inefficient and complex. The configuration of routers in the cloud model has almost always to be done manually and the routers are in many cases physically "one-armed". /13,14/

The concept of flow has emerged within the IP community over the past few years. A flow is defined as a sequence of packets sent from a particular host to a particular destination. These packets are related in terms of their routing and any local handling policy they may require. IP itself is connectionless, but many applications above IP are connection oriented. Efficient mapping of IP into ATM should consider the characteristics of the application or the transport protocol. Flows carrying real-time traffic, flows with a certain quality of service requirements or flows witch a long holding time will be handled most efficiently by mapping them into a individual ATM connection instead of hop-by-hop packet forwarding. Short duration flows would be best handled with the traditional hop-by-hop packet forwarding between IP routers using pre-established ATM connections between them. /14/

IP Switching uses standard ATM hardware, but totally changes the control software of the switch in order to operate the switch in a connectionless manner. The result of this is a combination of a IP router and an ATM switch, which is called an IP Switch. The name refers to the principle, that allows packet flows to be switched and bypass the router, when the routing information has been cached in the switch. The ATM switch is controlled by an IP Switch Controller, which is a high-end processor running standard IP routing software with extensions allowing it to make use of the switching hardware. The control protocol used is called General Switch Management Protocol (GSMP). The IP Switch is presented in Figure 20. The extensions in the controller include also a simple flow management protocol (Ipsilon Flow Management Protocol, IFMP) to associate IP flows with ATM virtual channels and also a flow classifier to decide whether to switch a flow or route it in a hop-by-hop basis. /14/

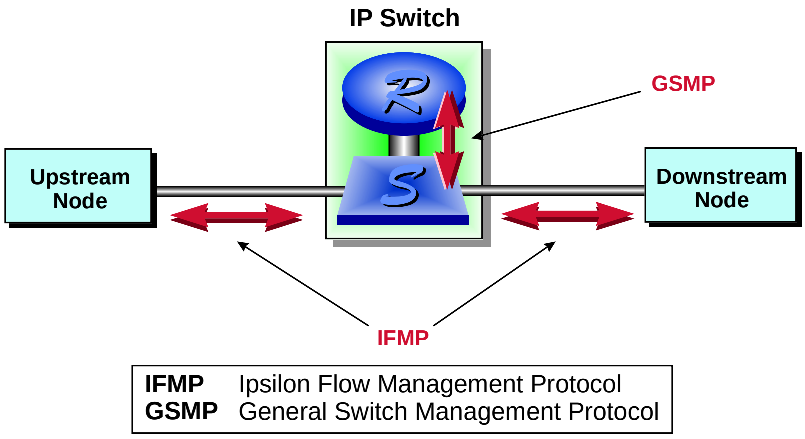

IP Switch
GSMP
R
S
Upstream Node
Downstream Node
IFMP
IFMP Ipsilon Flow Management Protocol
GSMP General Switch Management Protocol

Figure 20. IP Switch and Protocols.

At system start-up, a default ATM virtual channel is established between the IP Switch Controller and its neighbours. This VC is used for hop-by-hop connectionless forwarding of IP packets, also for long-term flows before the shortcut ATM VC is created. The IP packets are encapsulated using standard LLC/SNAP encapsulation as described in section 3.2.3. When the shortcut is created, the encapsulation used removes all IP header fields specified by the flow identifier. /14,15/

The process of deciding whether to switch or forward IP packets is discovered in the next section.

6.2 Identifying Flows and the Switching/Forwarding Decision

The IP flows are characterised according to the fields in the IP/TCP/UDP headers that determine the routing decisions (type of service, protocol, source and destination addresses, source and destination ports etc.). Two IP packets that have identical values of these fields, belong to the same flow. /14/

Ipsilon has specified three flow types, but currently only two flow types are used. The host-pair flow type is for traffic flowing between the same source and destination IP addresses. The port-pair flow type is for traffic flowing between the same source and destination IP addresses and also the same source and destination TCP/UDP port numbers. /14/

When the packet is first received across the default virtual channel it is reassembled and submitted to the IP Switch controller for forwarding. The controller forwards the packet in the normal manner and at the same time it performs a flow classification on the packet. According to this classification, the IP Switch Controller determines whether future packets belonging to the same flow should be switched directly or forwarded in the hop-by-hop basis by the controller. /14/

The Figure 21 describes the switching procedure.

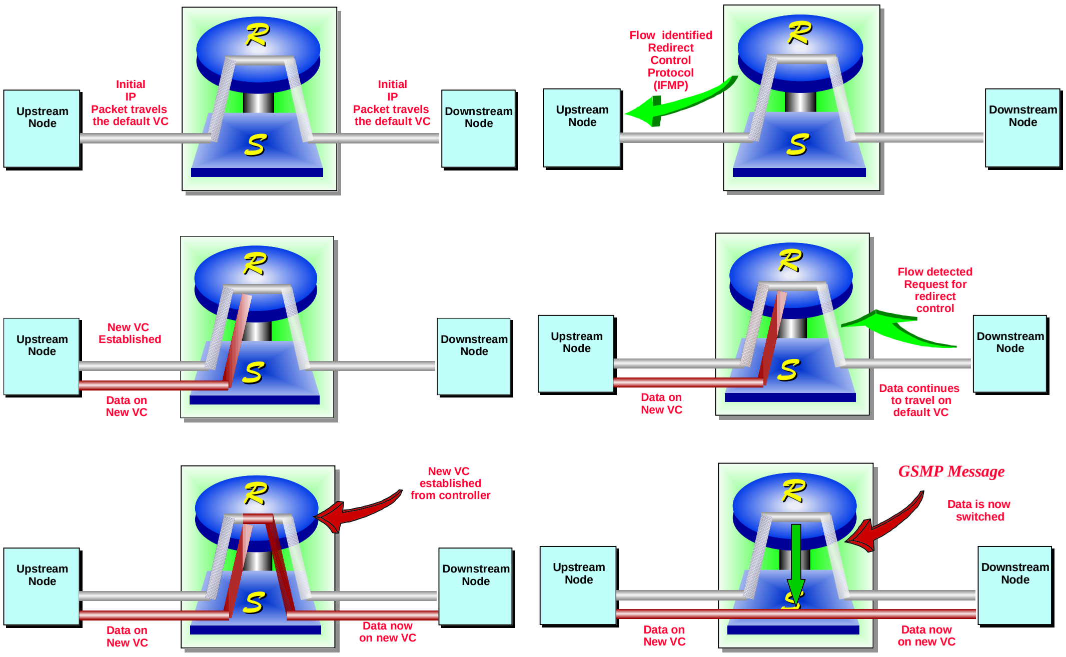


Figure 21. The Switching Procedure.

6.3 IP Switching Protocols

In developing IP Switching, Ipsilon has defined two protocols, the Ipsilon Flow Management Protocol and General Switch Management Protocol. They are published as IETF RFCs and are therefore publicly available. /14/

6.3.1 General Switch Management Protocol

The GSMP is a general purpose protocol to control the underlying ATM switch hardware. The GSMP is based on a master-slave architecture, where the controller acts as a master to the switch hardware. /14/

In the current implementation, the IP switch is composed of two separate components: ATM switch and a controller. The controller is connected to the ATM switch via a ATM link. The GSMP allows the controller to establish and release VCs across a switch, add and delete leaves on a point-to-multipoint connection, manage switch ports, request configuration information and request statistics. GSMP allows the switch to inform the controller of events, such as link down. The GSMP contains also an adjacency protocol, which is used to synchronise state across the link, to discover the identity at the other end of the link and to detect when it changes. /14/

6.3.2 Ipsilon Flow Management Protocol

The IFMP enables communications between IP devices (such as hosts) and IP Switches by associating IP flows with ATM VCs. The IFMP can be implemented in routers, LAN switches or IP hosts. /14/

The main function of IFMP is to instruct an adjacent node to attach a level 2 label to a specified IP flow. The IFMP defines the format for flow-redirect messages and acknowledgements. /14/

The first version of IFMP specification specifies three flow types and Ipsilon has implemented two of those at the moment. Flow type 1 is based on TCP/UDP source and destination addresses and port numbers. Flow type 2 is based on IP host addresses. The flow type 3 is based on IP network numbers, which would allow all traffic going between two sites to use a single circuit. /14/

6.4 IP Switching Networks

The current implementation of IP Switching includes two components. The main component is the IP Switch, which actually consists of a ATM switch and a IP Switch Controller. The functions of the IP Switch were described before. The other component of IP Switching is called IP Switch Gateway. It is actually an IP router that has extensions to support IFMP protocol. It connects Ethernet attached IP devices to the IP Switching cloud. There are also ATM NIC cards that support IFMP, so hosts can also be connected directly to IP Switching network. An example of an IP Switching network is showed in Figure 22. /10/

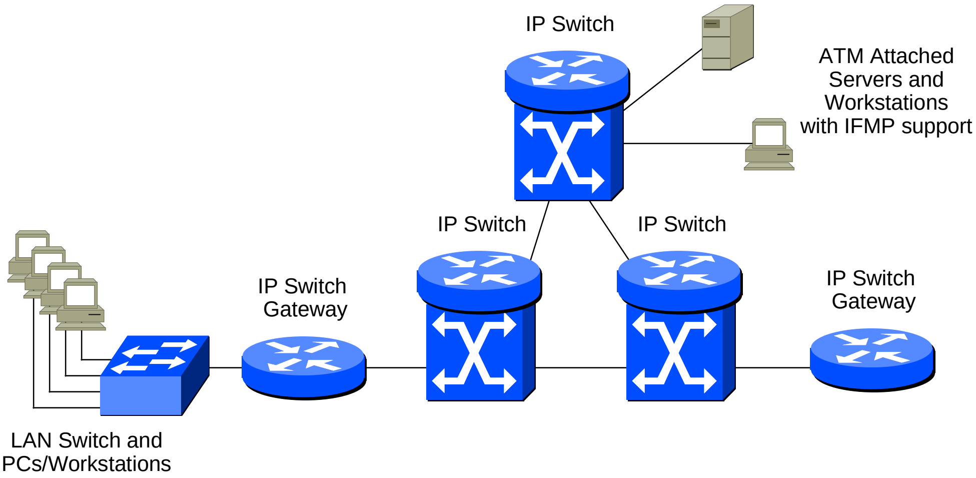


Figure 22. IP Switching Network Architecture.

The Ipsilon implementation currently support only two unicast routing protocols: OSPF and RIP. The multicast support comes with DVMRP and IGMP. How multicasting is done in IP Switching networks, is described in the following section. /10,14/

7 Multicasting in IP Switching Networks

IP Switching offers a very suitable and natural environment for standards based IP multicasting. IP Switches and the IP Switch Gateways (or end stations supporting IP Switching) form a typical IP based WAN network. In a typical IP based WAN, all point-to-point links between routers have their own IP subnets, that are subnetted with suitable subnet masks (for example 255.255.255.252). This kind of subnetting allows two hosts to each subnet. In IP Switching, the default VCs between IP Switches and IP Switch Gateways form similar point-to-point links between them. There are IP subnets for each links similar to normal WANs. The principle of addressing is shown in Figure 23. /14/

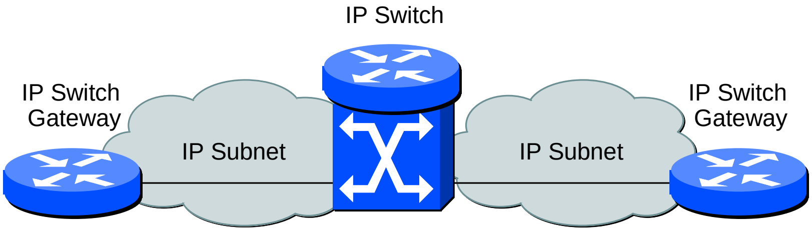


Figure 23. IP Addresses in IP Switching Network.

With that point-to-point oriented network infrastructure, the IP switches can support IP multicasting without modifications to standard IP multicasting protocols. The flow redirection mechanism is exactly the same as in unicast IP traffic. The exact functions of multicasting in IP switching networks are presented in this section. /14/

7.1 Multicasting - Short-Lived Traffic

With short-lived traffic, IP multicasting in IP Switching network is quite simple. The IP Switches change multicast routing and group membership information by using the standard DVMRP and IGMP messages. Using that information the IP Switch replicates the incoming multicast IP packets to all desired destinations via the default VC. The destination can then replicate those packets to all other destinations that it has. This procedure is done by the IP Switch Controllers and it is presented in Figure 24. /14/

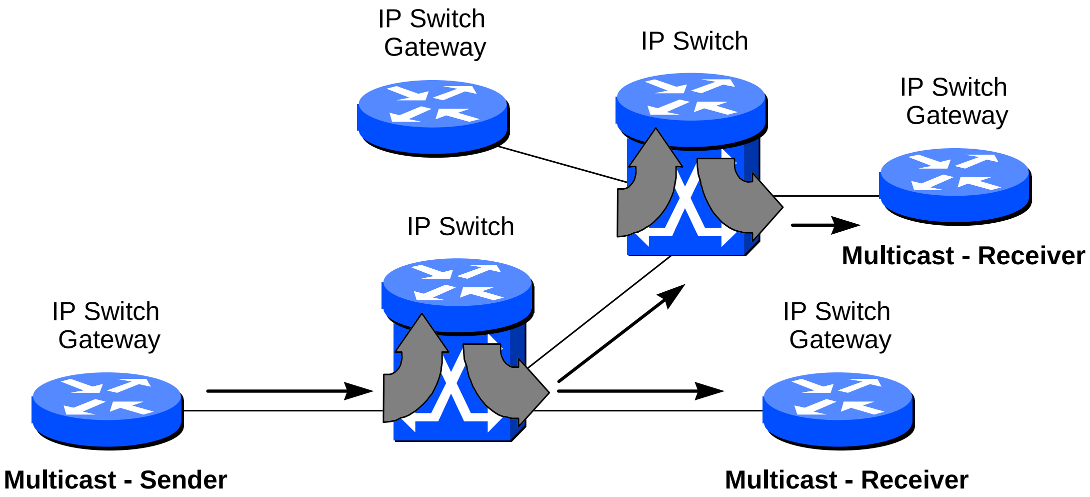


Figure 24. Multicasting with Short-Lived Traffic.

The IP multicasting with short lived traffic does not differ much from the PVC based native ATM service. It is quite processor intensive, because the IP Switch Controller does the replication. However, with longer lived flows another approach is used. /10/

7.2 Multicasting - Flow-Oriented Traffic

The previous section described IP multicasting with short-lived traffic. If the incoming IP multicast traffic can be identified as a flow, then the hardware multicast capability of the ATM switch may be used to offer better performance. The multicast capability of the ATM switch refers to point-to-multipoint VCs as described in section 5.2.2. /14/

The multicast group membership and routing information is learned via that same protocols as in multicasting of the short-lived IP traffic. The flow identification and redirection proceeds in exactly the same manner as for unicast traffic. The IP Switches and IP Switch Gateways recognise the incoming multicast flow and the flow is labelled. Then they redirect the flow onto a specific VC, but the VC in this case is point-to-multipoint VC. The root of the VC is the sending IP Switch Gateway (or an IFMP host) and then there are as many leaves as necessary to reach all branches. This is showed in Figure 25. /14/

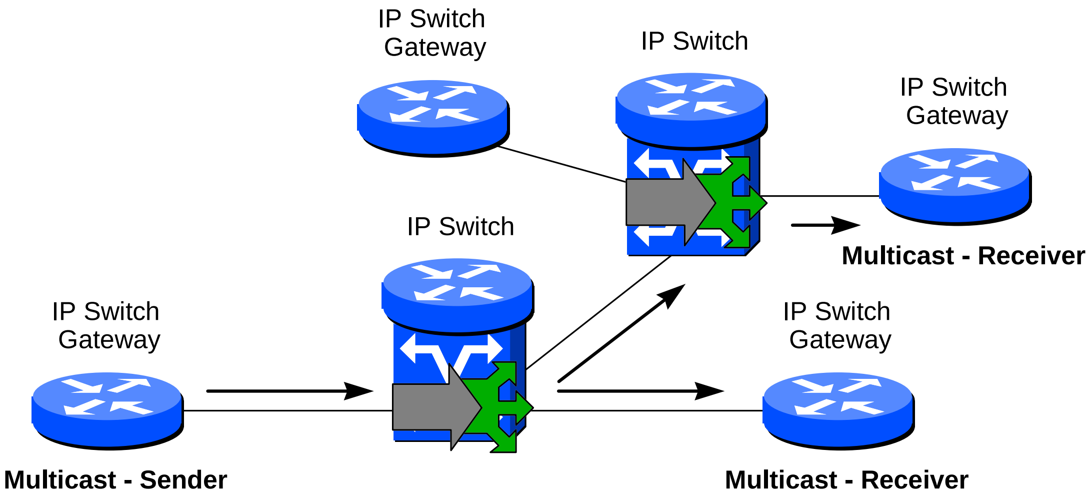


Figure 25. Multicasting with Flow-Oriented Traffic.

If the multicasting group membership protocol IGMP finds that there are new leaves to join the multicasting tree, the IP Switch Controller sends a GSMP message to the switch to add a branch to the point-to-multipoint ATM VC. /16,17/

The switch can also send a copy of the multicast flow to the IP Switch Controller, which can then forward the traffic to the default VC to reach also those branches that have not decided to redirect the flow to the shortcut ATM VC. /14/

7.3 Comparison to ATM Forum Technologies

This section will shortly compare multicasting in IP Switching environment to ATM Forum standard techniques.

As described in section 3.2.4, LAN Emulation operates at the layer two of the OSI reference model. LAN Emulation uses specialised server to implement the multicast functions. The server is called the Broadcast and Unknown Server (BUS). All LAN Emulation clients have unidirectional VCs (multicast send VC) to send multicast data to the BUS. The BUS will then forward all multicast data to all LECs that want to receive it (multicast forward VC). The BUS will therefore be heavily loaded if the multicast traffic is in an important role. Except forwarding, the BUS will also have other functions, such as buffering. It is also a single point of failure in the network. /9,18/

The ATM Forum have specified also a higher layer solution for carrying IP traffic over ATM networks. This is called the Multiprotocol Over ATM as described in section 3.2.4. Also the MPOA uses LANE multicasting mechanisms and therefore the same disadvantages apply. /10/

When compared to these, IP Switching look quite promising. The IP Switching infrastructure provides a much more natural approach to multicasting than the ATM

Forum solutions. There is no need for a separate multicast server and the existing standards multicast protocols can be used without any modifications.

8 Conclusions

There are multiple technologies for carrying IP traffic over ATM networks at the moment. Some of these are quite new and are based on proprietary solutions, although they may support standard features, too. IP Switching belongs to this category. /10/

The amount of IP traffic is growing rapidly and the applications are also developing all the time. The multimedia IP traffic is setting new challenges to the networks. Multicasting is one of these features. The previous sections described, how the IP Switching handled multicasting. IP Switching provides a promising alternative to the ATM Forum standard solutions. /10,18/

It is quite obvious that, Quality of Service (QoS) requirements are coming in all network technologies. Ipsilon's IP Switching provides a fairly suitable environment for these needs, too. The flow classification scheme provides a good possibilities to support QoS. As described earlier, the IP Switches forward the traffic either through the default VC or through the shortcut VC. Setting QoS for individual IP packets carried through the default VC is very difficult, but for the shortcut VCs it is quite natural. Each IP flow could be set up with some QoS requirements in the VC setup phase without the end user applications or protocols to be rewritten. /10,14/

The IP world will use a protocol called RSVP for QoS reservations. In other words the IP end stations use that for QoS needs. These requirements need to be mapped to ATM parameters, if the traffic is carried through the ATM network. How these mapping should be done in a suitable way is a well discussed matter. /10/

As was mentioned earlier, there are also other non-standard solutions similar to IP Switching from several companies (such as Cisco Systems, IBM, Cascade). In addition the ATM Forum is also setting its own standards. It will be interesting to see which solution will win the race or will there be something totally new.

References

/1/ Comer, Douglas E. Internetworking with TCP/IP, Vol I: Principles, Protocols, and Architecture, 3rd Edition. USA. Prentice-Hall, Inc. 1995. 613 p.

/2/ Internetworking Technology Overview. USA. Cisco Systems, Inc. 1994.

/3/ Chen, Thomas M & Liu, Stephen S. ATM Switching Systems. USA. Artech House Inc. 1995. 261 p.

/4/ I.150. B-ISDN Asynchronous Transfer Mode Functional Characteristics. International Telecommunications Union. 1993. 8 p.

/5/ Kyas, Othmar. ATM Networks. UK. International Thomson Computer Press Press. 1995. 372 p.

/6/ I.321. B-ISDN Protocol Reference Model and Its Application. International Telecommunications Union. 1991. 7 p.

/7/ Cisco Connection Documentation, Technology Information. Cisco Systems, Inc. 1997.

/8/ I.363. B-ISDN ATM Adaptation Layer (AAL) Specification. International Telecommunications Union. 1991. 96 p.

/9/ Alles, Anthony. ATM Internetworking. USA. Cisco Systems, Inc. 1995. 58 p.

/10/ Petrosky, Mary. Network Strategy Report: Shortcut Routing. USA. The Burton Group. 1997. 56 p.

/12/ Internetwork Design Guide. USA. Cisco Systems, Inc. 199X.

/13/ IP Switching: The Intelligence of Routing, The Performance of Switching. USA. Ipsilon Networks, Inc. 1996. 11 p.

/14/ Newman, Peter & Minshall, Greg & Lyon, Tom. IP Switching: ATM under IP. USA. Ipsilon Networks Inc. 1996. 14 p.

/15/ RFC1954. Transmission of Flow Labelled IPv4 on ATM Data Links. IETF. 1996.

/16/ RFC1953. Ipsilon Flow Management Protocol Specification for IPv4 Version 1.0. IETF. 1996.

/17/ RFC1987. Ipsilon's General Switch Management Protocol Specification Version 1.1. IETF. 1996.

/18/ Ilvesmäki, Mika. ATM-tekniikan käyttö internet-liikenteen välityksessä. TKK. 1996. 73 p.