NCS TIB 99-6

NATIONAL COMMUNICATIONS SYSTEM

TECHNICAL INFORMATION BULLETIN 99-6

MULTICAST OF INTERNET PROTOCOL PACKETS OVER ASYNCHRONOUS TRANSFER MODE NETWORKS

JUNE 1999

OFFICE OF THE MANAGER
NATIONAL COMMUNICATIONS SYSTEM
701 SOUTH COURT HOUSE ROAD
ARLINGTON, VA 22204-2198

NCS TECHNICAL INFORMATION BULLETIN 99-6

MULTICAST OF INTERNET PROTOCOL PACKETS OVER ASYNCHRONOUS TRANSFER MODE NETWORKS

JUNE 1999

PROJECT OFFICER:

DALE BARR, Jr.
Senior Electronics Engineer
Technology and Standards
Division

APPROVED FOR PUBLICATION:

PETER FONASH
Chief, Technology
and Standards Division

FOREWORD

Among the responsibilities assigned to the Office of the Manager, National Communications System, is the management of the Federal Telecommunications Standards Program. Under this program, the NCS, with the assistance of the Federal Telecommunications Standards Committee identifies, develops, and coordinates proposed Federal Standards which either contribute to the interoperability of functionally similar Federal telecommunications systems or to the achievement of a compatible and efficient interface between computer and telecommunications systems. In developing and coordinating these standards, a considerable amount of effort is expended in initiating and pursuing joint standards development efforts with appropriate technical committees of the International Organization for Standardization, the International Telecommunications Union-Telecommunications Standardization Sector, and the American National Standards Institute. This Technical Information Bulletin presents the results of an examination from an NS/EP perspective of selected technical interface considerations associated with the multicast of Internet Protocol (IP) packets over Asynchronous Transfer Mode (ATM) networks. Comments or statements of requirements which may assist in the advancement of this work are solicited and should be forwarded to:

Office of the Manager
National Communications System
Attn: Technology and Standards Division (N6)
701 S. Court House Road
Arlington, VA 22204-2198

This document was prepared under contract to the

Office of the Manager
National Communications System

Contract #DCA100-95-C-0126

by

SETA Corporation
6862 Elm Street
McLean, VA 22101

(703) 821-8178
Fax: (703) 821-8274

ABSTRACT

Internet Protocol (IP) multicast, a one-to-many transmission of IP datagrams via an internet using a single IP address, promises considerable cost savings in network and server resources when used to support group-based, distributed applications such as many of the evolving National Security and Emergency Preparedness (NS/EP) applications. However, IP multicast depends on the existence of a reliable underlying delivery system to forward data from senders to intended receivers. Data traversing the current Internet receives best-effort service only, and delivery is not guaranteed. Asynchronous transfer mode (ATM) is receiving broad acceptance as the base technology of the next generation of global broadband communications networks. As the Internet continues its evolution into a network of networks that supports better than best-effort traffic, the low delay and end-to-end quality of service (QoS) guarantees that ATM offers and the increased productivity and resource savings provided by multicast operations are expected to play a significant role. Resolution of the basic issue of mapping connectionless IP multicast service onto a non-broadcast multiple access media such as ATM is at the center of the ongoing work of the Internet Engineering Task Force (IETF) IP Over ATM (IPATM) Working Group and the ATM Forum's Multiprotocol Over ATM (MPOA) Working Group. This report presents the results of an examination from an NS/EP perspective of selected technical interface considerations associated with the multicast of IP packets over ATM networks. Specific issues examined in this report include address registration, address resolution, encapsulation, connection establishment, and routing.

TABLE OF CONTENTS

Section **Page**

EXECUTIVE SUMMARY ES-1

1.0 - INTRODUCTION 1
1.1 PURPOSE 1
1.2 SCOPE 1
1.3 BACKGROUND 1
1.4 ORGANIZATION 2
1.5 REVISIONS 3

2.0 - INTERNET PROTOCOL MULTICASTING/ATM ARCHITECTURE 4
2.1 ISO/OSI AND TCP/IP REFERENCE MODELS. 5
2.2 IP MULTICAST ROUTING 6
2.2.1 Dense Mode Approach 8
2.2.1.1 DVMRP 8
2.2.1.2 MOSPF 9
2.2.1.3 PIM-DM 10
2.2.2 Sparse Mode Approach 11
2.2.2.1 CBT 11
2.2.2.2 PIM-SM 11
2.2.3 Multicast Backbone (MBone) Interim Routing Approach 12
2.3 IP MULTICAST DELIVERY 13
2.3.1 RTP 14
2.3.2 RTCP 15
2.3.3 RSVP 15
2.3.4 RTSP 16
2.4 ASYNCHRONOUS TRANSFER MODE 17
2.4.1 ATM Multilayer Architecture 18
2.4.2 ATM Adaptation Layer 18

3.0 - ISSUES: IP MULTICASTING OVER ATM NETWORKS 20
3.1 ADDRESS REGISTRATION 20
3.2 ADDRESS RESOLUTION 23
3.3 ENCAPSULATION 25
3.4 CONNECTION ESTABLISHMENT 28
3.5 ROUTING 31

LIST OF REFERENCES 33

ACRONYMS 35

LIST OF FIGURES

Figure		Page
2-1	Simplified IP Multicast Model	4
2-2	The OSI Reference Model	5
2-3	Comparison of OSI and TCP/IP Models	6
2-4	Multicast Spanning Tree	7
2-5	DVMRP Routing	9
2-6	Nominal Core-Based Tree	12
2-7	IP Multicast Tunneling	13
2-8	ATM Functional Layering	18
	IP Multicast Over ATM	21
3-2	ATM Transport Network	26
3-3	Connection Establishment Message Flow	29

EXECUTIVE SUMMARY

PURPOSE

This report presents the results of an examination from a National Security and Emergency Preparedness (NS/EP) perspective of selected technical interface considerations associated with the multicast of Internet Protocol (IP) packets over Asynchronous Transfer Mode (ATM) networks. Specific issues examined in this report include address registration, address resolution, encapsulation, connection establishment, and routing.

BACKGROUND

Recent advances in a wide variety of communications and information dissemination applications will improve considerably the ability of planners and providers of NS/EP services to communicate and collaborate during NS/EP operations. Among the advances of relevance to the NS/EP community are new or improved multimedia and real-time interactive applications in the fields of video teleconferencing, collaborative consultation, distributed parallel processing, telescience, visualization, and distributed simulation. These applications not only enhance communications and collaboration, but also promise to leverage considerably more value from supporting telecommunications networks with minimal increase in the level of network investment. Many of the cited applications are bandwidth-intensive, real-time, group-based, distributed programs, frequently involving widely dispersed locations. As such, they generally require, high-bandwidth, simultaneous communications from each data source to multiple destinations, often over a wide area network (WAN). They also generally require scalable protocols capable of providing guaranteed real-time quality of service (QoS) on an end-to-end basis. QoS refers to the ability to ensure that packet flow through the network is sustained at an agreed-on throughput and that some types of packets are able to receive preferential treatment. Because of cost savings in network and server resources, IP multicast, a technology which uses the Internet to send and receive information from a group of hosts using a single transmit operation, is receiving considerable attention for support of group-based, distributed applications. IP multicast depends on a reliable underlying delivery system to forward data from senders to intended receivers. However, data traversing the current Internet receives best-effort service only, and delivery is not guaranteed. Additionally, not all routers on the Internet are multicast-enabled. Because of this and other factors such as bit errors during transmission, variable router queuing delays, and loss due to network congestion, the current Internet packet delivery infrastructure lacks the capability to meet the stringent IP multicast QoS requirements of some of the potential NS/EP applications cited above.

ATM is receiving widespread acceptance as the base technology of the next generation of global broadband communications networks. This widespread acceptance is due in large measure to the

ability of ATM networks to (1) efficiently interwork ATM with local area network (LAN) and WAN technologies, (2) guarantee the QoS delivered across ATM backbones, and (3) provide dedicated, high performance, high-speed connections between users. Because of these capabilities, large segments of the telecommunications research, development, and standards communities are interested in expanding the current Internet architecture to support real-time multicasting of IP packets over ATM networks. Both the Internet Engineering Task Force's (IETF's) IP Over ATM (IPATM) Working Group and the ATM Forum's Multiprotocol Over ATM (MPOA) Working Group are heavily involved in coordinating IP multicast over ATM standardization activities. As the Internet continues its evolution into a network of networks that supports better than best-effort traffic, the low delay and end-to-end QoS guarantees that ATM offers and the increased productivity and resource savings provided by multicast operations are expected to play a significant role. Resolution of the basic issue of mapping connectionless IP multicast service onto a non-broadcast multiple access media such as ATM is a central focus of the ongoing work of the IETF's IPATM and ATM Forum's MPOA Working Groups.

IP MULTICASTING

IP datagrams are the fundamental packets of information in the Internet. IP datagrams are logical groupings of information sent as network layer units over the transmission medium. In this document, the term datagram is used synonymously with packet. Rather than sending one datagram to each destination, IP multicast sends one datagram to a multicast group identified by a single IP destination address in the IP datagram header. The datagram is then replicated, as required by the network's multicast routers and switches, to enable all members of the multicast group to receive the broadcast (even in scattered subnetworks). Each datagram is independent and has no relationship with other datagrams. Routing and functions are performed by IP routers using multicast routing protocols. Source and destination addresses are derived from the IP address. Each address consists of a network identifier, an optional subnetwork identifier, and a host identifier. Datagrams are forwarded by the network based upon the network identifier portion of the address. Receivers wishing to subscribe to an IP multicast group inform their local routers and join the group.

IP MULTICAST ROUTING

In IP multicast networks, the network must be able to build packet distribution trees that specify a unique forwarding path between the subnet of the source to each subnet containing members of the multicast group. One of the key components is the router. Routers use forwarding algorithms of one type or another to forward packets from a source to a specified group of receivers. One algorithm used is the spanning tree. A spanning tree is an algorithm used to create a logical topology that connects all network segments and ensures that only one path exists between any two nodes. Forwarding along the branches of a spanning tree guarantees that the

multicast packet will not loop and that it will eventually reach all routers in the network. Several algorithmic standards are available for building multicast spanning trees. Depending on the expected distribution of multicast group members throughout the network and the availability of bandwidth, all of the algorithms generally follow one or two basic approaches. The first approach, the "dense mode" approach, is based on the assumption that multicast group members are densely distributed throughout the network and that bandwidth is plentiful. The second approach, the "sparse mode" approach, basically assumes that multicast group members are sparsely distributed throughout the network and bandwidth is not necessarily widely available. Common to both approaches is the need to set up a state in intermediate routers for multicast forwarding. The approaches differ mainly on who initiates the state creation) the sender, the receiver, or the routers themselves. The principal dense mode multicast routing protocols are the Distance Vector Multicast Routing Protocol (DVMRP), Multicast Open Shortest Path First (MOSPF), and Protocol Independent Multicast-Dense Mode (PIM-DM). The principal sparse mode multicast routing protocols are the Core-based Trees (CBT) and PIM-Sparse Mode (PIM-SM) protocols.

MULTICAST BACKBONE INTERIM ROUTING APPROACH

Relatively few routers in the Internet currently support IP multicasting. However, as old routers reach the end of their programed lives, most of the new replacement routers do tend to be multicast capable. Several years ago, researchers seeking a way to enable the deployment of a resource and testbed for testing multicasting protocols and applications, developed the Internet Multicast Backbone (MBone) as an interim solution. The MBone is an experimental, cooperative volunteer effort spanning several continents. It is a virtual network layered on top of the Internet created by an interconnected set of subnetworks and routers. The MBone supports routing multicast packets without disturbing or altering other Internet traffic. It originated from experiments conducted during IETF meetings in which live audio and video were transmitted around the world, and has been in existence since early 1992. MBone uses a network of routers called "MRouters" that can support IP multicast, using augmented "tunnels" that forward multicast packets between islands of MBone subnets. Tunneling is the practice of encapsulating a message from one protocol in another protocol and using the second protocol to traverse network hops, or connect islands of multicast routes that are separated by links that do not support IP multicast. Most routers used on the Internet today are unicast, (i.e., point-to-point) routers. MBone multicast traffic bypasses the unicast routers on the Internet by the use of software that encapsulates the multicast packets in traditional unicast packets so that unicast routers can handle the information. At the destination, the encapsulation is stripped off and the original message is reintroduced to the network at its destination.

IP MULTICAST DELIVERY

The Transmission Control Protocol/Internet Protocol (TCP/IP) used on the Internet was designed primarily for the reliable transmission of unicast data with minimal or no delay constraints. TCP works well in this context. However, the Internet, like other packet networks, sometimes loses and reorders packets and delays them by variable amounts of time. This fact, and its inability to provide end-to-end QoS guarantees, generally makes the current Internet unreliable for the transmission of multimedia traffic. Multimedia traffic is expected to comprise a significant portion of potential NS/EP multicast traffic. Multimedia traffic exhibits different response characteristics with respect to delay in the delivery of datagrams, and has considerably more stringent QoS requirements. Because of these requirements, TCP is generally not adequate to support real-time multimedia multicast applications. Consequently, additional protocols are required to provide the necessary transport services. The Internet community) IETF working groups and industry vendors) is working to develop reliable multicast protocols to overcome the limitations cited above. These protocols are currently at various levels of maturity. The principal protocols currently under development include the Real-Time Transport Protocol (RTP), Real-Time Control Protocol (RTCP), ReSerVation Protocol (RSVP), and Real-Time Streaming Protocol (RTSP).

ASYNCHRONOUS TRANSFER MODE

ATM is a cell-based transfer mode that requires variable length user information of multiple types (e.g., voice, video, or data) to be segmented and reassembled to and from short, fixed length cells. Fixed length 53-byte cells allow cell processing to occur in hardware, thereby reducing transit delay. The first 5 bytes of the cell contain cell-header information, and the remaining 48 bytes contain the payload, or user information. ATM networks are fundamentally connection-oriented. That is, they require that a virtual circuit/channel (VC) be setup across the ATM network prior to the transfer of any data. A VC is a communications channel that provides for the unidirectional transport of ATM cells. ATM circuits are of two types: virtual paths (VPs), identified by VP identifiers (VPIs); and virtual channels, identified by the combination of a VPI and a VC identifier (VCI). VPIs and VCIs are unique numerical tags defined by fields in the ATM cell header. A VP is a bundle of VCs, all of which are switched transparently across the ATM network on the basis of the common VPI. However, all VCIs and VPIs have only local significance across a particular link and are remapped, as appropriate, at each switch.

The ATM multilayer architecture uses a logical model to describe the functionality it supports. ATM functionality corresponds to the physical layer and part of the data link layer of the Open Systems Interconnection (OSI) reference model. The ATM architecture consists of three functional layers) the physical layer, the ATM layer, and the ATM adaptation layer (AAL). The physical layer is the lowest layer. The physical layer is responsible for the transmission of cells between two ATM hosts across a specific physical link. The ATM layer is the next layer above

the physical layer. It is independent of the physical layer and the AAL above it. The ATM layer is responsible for a number of functions concerning the cell header, including cell header generation and extraction. Other functions performed by the ATM layer are cell multiplexing and demultiplexing, and traffic and congestion control. The last layer, the AAL, is the fundamental layer for encapsulation operations. AAL sits on top of the ATM layer. Its primary purpose is to adapt the flow of information received from a higher layer application like voice or data from IP or other upper layers to the ATM layer.

ISSUES: IP MULTICASTING OVER ATM NETWORKS

The IETF's IPATM Working Group Charter states its intention to "...focus on the issues involved in running internetworking protocols over ATM networks." According to the charter, "The final goal of the Working Group is to produce standards for the TCP/IP protocol suite and recommendations which could be used by other internetworking protocol standards...". [1] There are significant differences in the way IP and ATM operate. The principal difference involves the nature of the two technologies, particularly the methods used for handling the transport layer services of connection management, data transfer, and flow control. IP provides a connectionless service which allows the transfer of information divided into packets, or datagrams, among service subscribers without the need for end-to-end establishment of a transmission path. Each datagram is transmitted individually and can even follow different routes to its destination. Once all the datagrams forming a message arrive at the destination, they are recompiled into the original message. Once a VC is established, the ATM protocol selects a physical route from source to destination and enters the information in a route table in the switch. All switches along the pathway make table entries so they can route packets accordingly. All messages for that connection follow the same path to the destination. Use of the services of ATM for IP multicasting requires a mechanism which basically makes the difference in characteristics between ATM and IP transparent to the ATM network. Achieving transparency involves significant issues of address registration, address resolution, encapsulation, connection establishment, and routing.

ADDRESS REGISTRATION

Each device within a network must have a unique and identifiable address in order to receive and transmit messages. To establish an ATM connection at the user-network interface (UNI), both the user and the network must know the ATM addresses in effect at that UNI. These addresses are used in the Calling Party Number information elements of signaling messages sent by the user, and in Called Party Number information elements of signaling messages sent to the user. Address registration is the dynamic exchange of network routing prefixes on the network side and end system identifiers on the host side. Existing network layer protocols, (e.g., IP) have their own addressing schemes and associated routing protocols. However, in an integrated IP over ATM environment, the ATM layer is decoupled from the IP protocol and is defined by its own

addressing structure. The ATM addressing space is logically disjointed from the addressing space of whatever protocol runs over it and typically would not bear any relationship to it. Hence, all protocols operating over ATM require some form of address registration and the use of an ATM address resolution protocol to map higher layer addresses to their corresponding ATM addresses. The Internet Information Center (InterNIC) is the combined name for the providers of IP address registration, information, and database services to the Internet. With ATM, there are several registration authorities from which network managers can acquire unique network name and address space. In the United States, ATM code points may be obtained from the National Institute of Standards and Technology (NIST) and the American National Standards Institute (ANSI). To facilitate the administration and configuration of addressing in an integrated IP/ATM environment, the IETF Networking Working Group and the ATM Forum Technical Committee are currently studying address registration and address resolution issues.

ADDRESS RESOLUTION

Address resolution is used to resolve differences between different addressing schemes. Typically, address resolution specifies a technique for mapping network layer (Layer 3) addresses to data link layer (Layer 2) addresses. In a classical IP environment, the source and destination address fields of the IP datagram header contain the addressing information needed to route datagrams in a connectionless network. IP routers using static or dynamic lookup or routing tables attempt to match the network address contained in the header of a datagram with a network address entity contained in the routing table. If the destination node is on its local network, the datagram is forwarded directly to the destination host. If the destination node is on some other network, the datagram is forwarded to the IP local router for forwarding, as appropriate. The IP multicast model is a receiver-initiated model. Receivers wishing to subscribe to a multicast group use the Internet Group Management Protocol (IGMP) to inform their local router. Routers disseminate membership information to all other routers in the routing domain. An IP sender wishing to transmit data to a multicast group sends the IP packets to the IP address of the multicast group. Over ATM networks, a mechanism is required to map IP multicast group addresses to corresponding ATM addresses. The mechanism used is the Multicast Address Resolution Server (MARS). The MARS acts as a registry, associating IP Layer 3 multicast group addresses to one or more ATM interface addresses representing the group's members. Each ATM-based host and router client communicates with the MARS by using a globally known VC. The MARS may reside within any ATM endpoint that is directly addressable by the endpoints it is serving.

A potential concern with the ATM Forum's MPOA model multicast arrangement is that while MARS appears to be an effective approach for small environments, several studies have concluded that it will not scale well to large networks where multicast traffic must transit to a single multicast server. Other concerns are that (1) the current proposal for MARS does not appear to provide sufficient flexibility to handle the range of new application requirements for QoS and traffic behavior, and (2) the requirement to establish dedicated VCs between the root

and each recipient leaf node and management of the MARS mapping information requires considerable overhead. Discussion is currently ongoing within the IETF concerning ways to avoid the overhead associated with the establishment and maintenance of separate multicast VCs.

ENCAPSULATION

Encapsulation is an operation which allows a network to carry traffic using non-similar protocols through the transport network. The operation involves enclosing data formatted by protocols operating at the upper layers of the OSI layered model (e.g., IP datagrams or protocol data units [PDUs]), within another protocol that performs lower layer bearer services (e.g., an ATM cell), in order to transport the encapsulated data across a network for which the original protocol was not designed. Ideally, the transport network does not become involved with either the syntax or the format of the transported traffic. The function of mapping IP PDUs into the information field of the ATM cell and vice versa is performed in the AAL. When a VC is created, a specific AAL type is associated with the VC.

Because ATM is a cell-based transfer mode, variable IP multicast data must be formatted into short, fixed length cells prior to transport across the ATM network. IP packets are often considerably larger than ATM cells. Consequently, there is generally a requirement to reduce the size of the IP multicast packet to adapt it to the ATM cell size. As part of the encapsulation function, fragmentation (partitioning of IP datagrams into parts) is used to separate datagrams that are too large for the supporting transport network technology to support. In the transmit direction, encapsulation generates an appropriate cell header for the information field in the ATM layer, less the header error control (HEC), which is the responsibility of the physical layer. It may also include translation from a service access point (SAP) identifier to a VPI and VCI. An SAP is a physical interface between the layers of the OSI model through which lower layers provide services to the higher layers passing over the PDUs. In the receive direction, a decapsulation function is performed by the ATM layer that includes extracting the ATM cell header and passing the cell information field to the AAL. The AAL maps the information field contents into appropriate PDUs for forwarding to the upper layer protocols. Encapsulation introduces an element of delay which could have an impact on the response characteristics of multimedia traffic. Multimedia traffic is expected to comprise a significant portion of future potential NS/EP multicast traffic.

CONNECTION ESTABLISHMENT

Since ATM is connection-oriented, a connection request needs to be routed from the requesting node through the ATM network to the destination node, much as packets are routed within a packet-switched network. Connection establishment for switched virtual connections/circuits (SVCs) (connections established via signaling) is by mutual agreement and can be set up using a simple set of user commands. For unicast connections, the signaling protocol used is an

exchange of messages between the caller and receiver across an adjacent ATM switch. Multicast connections are implemented in a slightly different manner than unicast connections. Multicast connections are supported by a collection of sender-initiated point-to-multipoint, unidirectional connections and associated endpoints. This arrangement requires the router to know each intended recipient and explicitly establish a connection between itself as the root and each recipient as a leaf node. Leaf nodes may be added or dropped at any time after establishing the connection. The calling party/host sends the frames to the MARS and not to the client, by using the point-to-point connection cited above. During address resolution, the ATM address of the MARS is provided and not the address of the end user. Information flowing from the source is replicated at the router. Three methods have been proposed for the flow of information onward to the end-user) the VC mesh, the Multicast Server (MCS), and VP multicasting. In the VC mesh method, separate point-to-point circuits connect the MARS to all end-users/clusters of members. The MCS method establishes point-to-multipoint connections to the final destination end-users. In turn, end-users wishing to receive multicast traffic, need only connect to the MCS. VP multicasting, a third connection method under discussion in certain forums, would provide multipoint-to-multipoint VP connections to link all nodes in a multicast group, with each node given a unique VCI value within the VP. Interleaved packets could then be identified by the unique VCI value of the source. However, this mechanism would require a protocol to uniquely allocate VCI values to nodes, and at present no such protocol exists. In addition, it is also not certain whether current segmentation and reassembly (SAR) devices could easily support such a mode of operations. It appears that connections through both a VC mesh and an MCS could increase packet delay due to the setup time of lengthy connection establishment procedures. However, because of the bi-directional nature of MCS connections, versus the unidirectional nature of mesh connections, it seems that delay would be higher using an MCS than a mesh.

ROUTING

Routing IP multicast packets across an ATM network is a topic of current discussion both in the ATM Forum and IETF networking groups. The two network entities represent two independent routing approaches and hierarchies that make it difficult to adequately coordinate routing across topologies. IP multicast routing is unaware and independent of the ATM topology. For a large IP network running over a large ATM network, this implies that the user needs to install and manage two independent routing hierarchies (i.e., one for ATM, and one for IP). To forward an IP multicast packet to destinations across an ATM network, the packets traverse the network of IP routers following the path specified by the standard routing computation, until they reach an IP router at the ATM network interface. At this node the routing function of the IP subnet is terminated. The router forwards the IP packet across an existing permanent virtual circuit (PVC) if available, or may buffer the packet while setting up an SVC to the associated address. The created SVC remains for a specified period of time and closes if no traffic is passed through it. This process can result in some delay in forwarding packets. Additionally, because of differences in approaches to specifying QoS routing in both topologies, a potential complication exists in the transport of IP multicast packets over ATM networks. Ideally when an IP host commits to

provide a specified type of service for an application, it must be able to request an appropriate QoS from the ATM network using the ATM service model. Although mechanisms exist today for traffic prioritization on router-based IP networks, they do not appear to be well suited to the demands of ATM networks. The IETF is examining several mechanisms by which QoS specifications for IP multicast can be translated into QoS specifications that are meaningful for an ATM network. The principal mechanisms include both receiver-initiated mechanisms, and depending on the distribution environment, receiver/sender-initiated mechanisms. However, in ATM networks, resource reservations are made at connection setup, using UNI and network-network interface (NNI) signaling protocols. The differences between receiver-initiated mechanisms and ATM state establishment could present potential problems in that the service priorities established at the IP subnet may not be carried through the ATM network, thereby creating network inefficiencies when executing IP service contracts.

SECTION 1.0

INTRODUCTION

1.1 PURPOSE

This report presents the results of an examination from a National Security and Emergency Preparedness (NS/EP) perspective of selected technical interface considerations associated with the multicast of Internet Protocol (IP) packets over Asynchronous Transfer Mode (ATM) networks.

1.2 SCOPE

This report contains: (1) an overview discussion of IP multicast operations and a brief introduction to the ATM multilayered architecture, and (2) an examination from an NS/EP perspective of selected technical issues related to the multicast of IP packets over ATM networks. Specific issues examined in this report include address registration, address resolution, encapsulation, connection establishment, and routing.

1.3 BACKGROUND

Recent advances in a wide variety of communications and information dissemination applications will improve considerably the ability of planners and providers of NS/EP services to communicate and collaborate during NS/EP operations. Among the advances of relevance to the NS/EP community are new or improved multimedia and real-time interactive applications in the fields of video teleconferencing, collaborative consultation, distributed parallel processing, telescience, visualization, and distributed simulation. These applications not only enhance communications and collaboration, but also promise to leverage considerably more value from supporting telecommunications networks with minimal increase in the level of network investment. Many of the cited applications are bandwidth-intensive, real-time, group-based, distributed programs, frequently involving widely dispersed locations. As such, they generally require, high-bandwidth, simultaneous communications from each data source to multiple destinations, often over a wide area network (WAN). They also generally require scalable protocols capable of providing guaranteed real-time quality of service (QoS) on an end-to-end basis. QoS refers to the ability to ensure that packet flow through the network is sustained at an agreed-on throughput and that some types of packets are able to receive preferential treatment. Because of cost savings in network and server resources, IP multicast, a technology which uses the Internet to send and

receive information from a group of hosts using a single transmit operation, is receiving considerable attention for support of group-based, distributed applications. IP multicast depends on the existence of a reliable underlying delivery system to forward data from senders to intended receivers. However, data traversing the current Internet receives best-effort service only, and delivery is not guaranteed. Additionally, not all routers on the Internet are multicast-enabled. Because of this and other factors such as bit errors during transmission, variable router queuing delays, and loss due to network congestion, the current Internet packet delivery infrastructure lacks the capability to meet the stringent QoS requirements of some of the potential NS/EP applications cited above.

ATM is receiving widespread acceptance as the base technology of the next generation of global broadband communications networks. This widespread acceptance is due in large measure to the ability of ATM networks to (1) efficiently interwork ATM with local area network (LAN) and WAN technologies; (2) guarantee the QoS delivered across ATM backbones; and (3) provide dedicated, high performance, high-speed connections between users. Because of these capabilities, large segments of the telecommunications research, development, and standards communities are interested in expanding the current Internet architecture to support real-time multicasting of IP packets over ATM networks. Both the Internet Engineering Task Force's (IETF's) IP Over ATM (IPATM) Working Group and the ATM Forum's Multiprotocol Over ATM (MPOA) Working Group are heavily involved in coordinating IP multicast over ATM standardization activities. As the Internet continues its evolution into a network of networks that supports better than best-effort traffic, the low delay and end-to-end QoS guarantees that ATM offers, and the increased productivity and resource savings provided by multicast operations are expected to play significant roles. Resolution of the basic issue of mapping connectionless IP multicast service onto a non-broadcast multiple access media such as ATM is a central focus of the ongoing work of the IETF's IPATM and ATM Forum's MPOA Working Groups.

1.4 ORGANIZATION

This document is further divided into the following subsequent sections:

- Section 2.0, *Internet Protocol Multicasting/ATM Architecture*, provides a brief overview discussion of the IP multicast model and the ATM multilayered architecture with emphasis on the ATM adaptation layer (AAL).

- Section 3.0, *Issues: IP Multicasting Over ATM Networks*, examines from an NS/EP perspective selected technical issues related to the multicast of IP packets over ATM networks.

1.5 REVISIONS

This document will be updated as directed by the Technology and Standards Division (N6), Office of the Manager, National Communications System (OMNCS). Comments and recommendations which may assist in the advancement of this effort are solicited and should be forwarded to:

Office of the Manager
National Communications System
Attn: N6
701 Court House Road
Arlington, VA 22204-2198

SECTION 2.0

INTERNET PROTOCOL MULTICASTING/ATM ARCHITECTURE

IP datagrams are the fundamental packets of information in the Internet. IP datagrams are logical groupings of information sent as network layer units over the transmission medium. In this document, the term datagram is used synonymously with packet. Rather than sending one datagram to each destination, IP multicast sends one datagram to a multicast group identified by a single IP destination address in the IP datagram header. The datagram is then replicated, as required by the network's multicast routers and switches, to enable all members of the multicast group to receive the broadcast (even in scattered subnetworks). Each datagram is independent and has no relationship with other datagrams. Routing functions are performed by IP routers using multicast routing protocols. Source and destination addresses are derived from the IP address. Each address consists of a network identifier, an optional subnetwork identifier, and a host identifier. Datagrams are forwarded by the network based upon the network identifier portion of the address. Receivers wishing to subscribe to an IP multicast group inform their local routers and join the group. Multicast groups may be permanent or transient. Permanent groups have well-known, administratively assigned IP addresses. However, it is the address, not the membership of the group, that is permanent. At any time permanent groups may have any number of members, even zero. Transient groups exist only as long as they have members. Hosts may join and leave groups at any time. The sender does not need to maintain a list of receivers. There are no restrictions on the location or number of members. A host may be a member of more than one group at a time. At the application level, multiple applications may share a single group address on a host. The flexibility inherent in the above arrangement makes changing membership relatively easy to handle even in large networks. Figure 2-1 shows a simplified depiction of an IP multicast model.

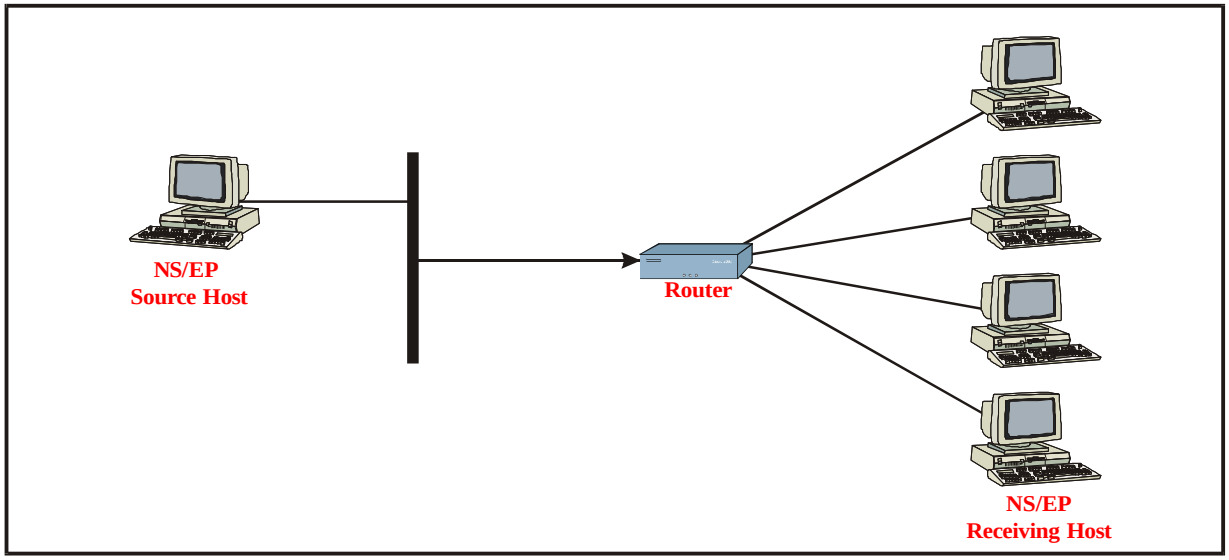

Figure 2-1. Simplified IP Multicast Model

2.1. ISO/OSI AND TCP/IP REFERENCE MODELS

The International Standards Organization (ISO)'s seven-layer Open System Interconnection (OSI) reference model and the Defense Advanced Research Projects Agency's (DARPA's) four-layer Transmission Control Protocol/Internet Protocol (TCP/IP) reference models provide standard architectures that define responsibility for communications tasks. In both models, each layer provides a function or group of functions. One or more entities implement functions at each layer. The entities within a layer interact directly with the layer immediately below it and provide services for use by the layer above it. The OSI reference model in particular is used universally as a method for teaching and understanding network functionality. Corresponding entities on either side of the OSI reference model communicate with each other by means of a common protocol (e.g., an entity or peer at the physical layer on system A communicates with its peer at the physical layer on system B on the other side). Figure 2-2 shows the seven layers of the OSI reference model. It is important to note that there is no direct communication between peer layers except at the physical layer. Above the physical layer, each protocol entity sends data down to the next lower layer, until it reaches the physical layer, then across and up to its peer on the other side. With connectionless packet service, even the physical layer may not be directly connected to its peer on the other side. However, peer layers must share a common protocol in order to communicate.

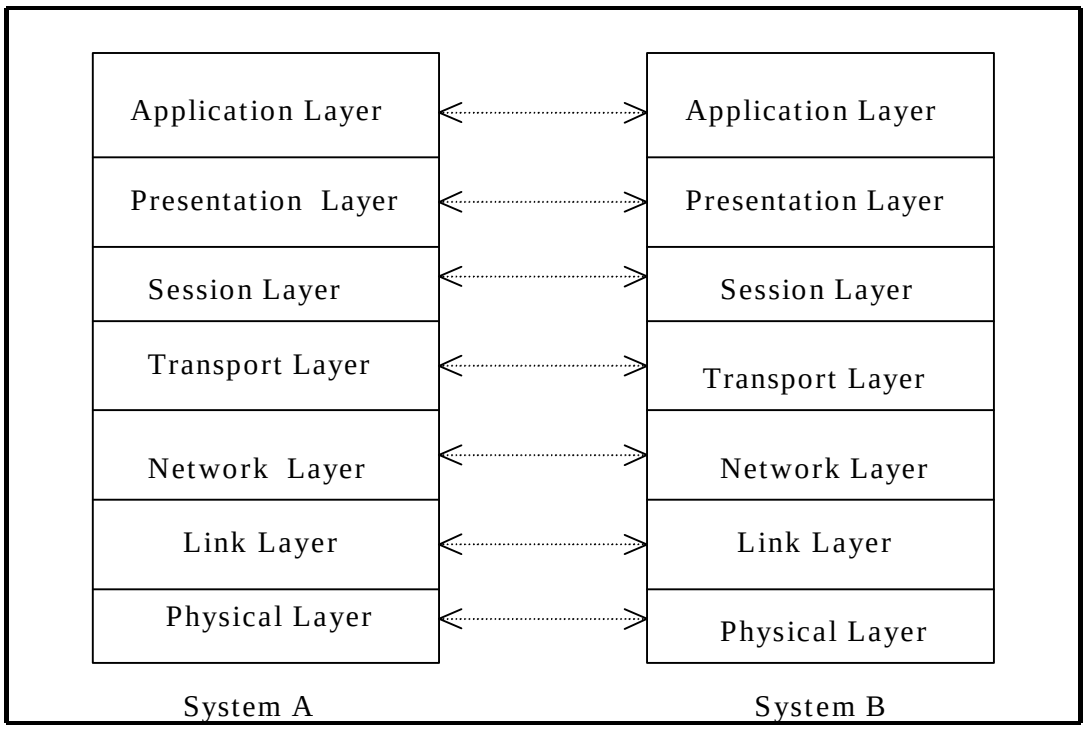

Figure 2-2. The OSI Reference Model

TCP/IP is actually a common name for a suite of protocols developed to support the construction of worldwide internetworks. TCP and IP are only the best-known protocols of the suite. The TCP/IP suite of protocols has been adopted by the Department of Defense (DoD) as its choice of network communications protocols. The TCP/IP model and the OSI reference model are both used to describe network protocol layers. However, there are very real differences between the two reference models. Figure 2-3 shows the relative functional positions of the OSI and TCP/IP layers. The modern Internet represents a fusion of both models. The TCP/IP protocol suite consists of the core Internet protocols. IP multicasting is a function of the TCP/IP protocol suite. Two of the most important TCP/IP multicasting functions are to (1) provide a mechanism to dependably route data to its proper destination, and (2) ensure that the data delivered reliably reflects the data transmitted.

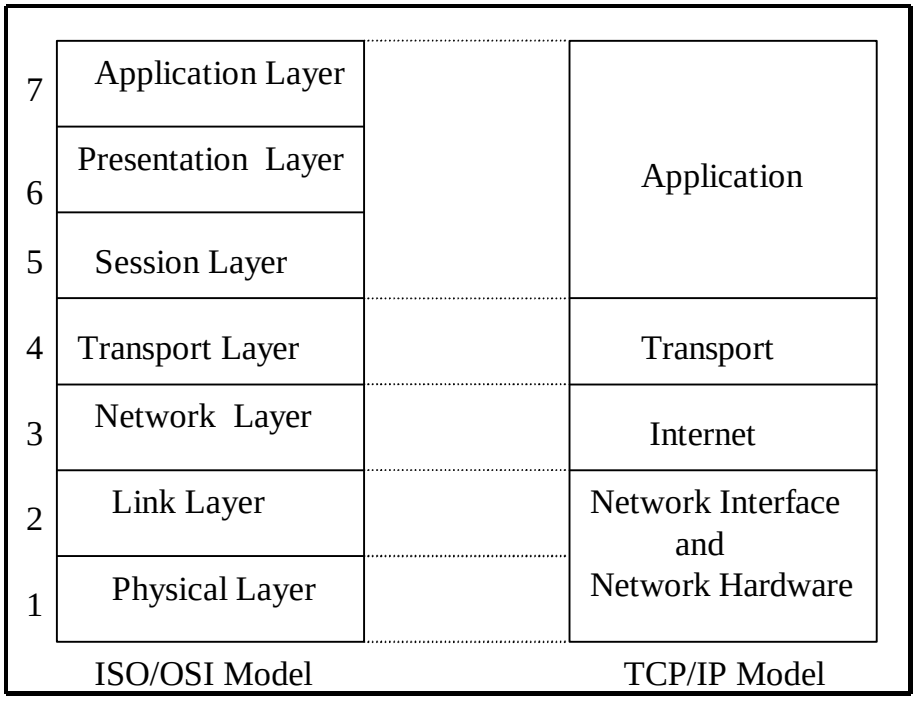

Figure 2-3. Comparison of OSI and TCP/IP Models

2.2. IP MULTICAST ROUTING

In IP multicast networks, the network must be able to build packet distribution trees that specify a unique forwarding path between the subnet of the source to each subnet containing members of the multicast group. One of the key components is the router. Routers use forwarding algorithms of one type or another to forward packets from a source to a specified group of receivers. One algorithm used is the spanning tree. A spanning tree is an algorithm used to create a logical topology that connects all network segments and ensures that only one path exists between any two nodes. The objective in constructing the spanning tree is to ensure that only one

copy of each packet is forwarded on each branch of the tree. Forwarding along the branches of a spanning tree guarantees that the multicast packet will not loop and that it will eventually reach all routers in the network. The method used to construct a spanning tree varies depending on the IP multicast routing protocols used. However, once a spanning tree is constructed, all multicast traffic is distributed over it. Figure 2-4 provides a simplified depiction of a spanning tree. The last router at the destination network determines the data-link address of the recipient and forwards the datagram directly to the host. Several algorithmic standards are available for building multicast spanning trees. Depending on the expected distribution of multicast group members throughout the network and the availability of bandwidth, all of the algorithms generally follow one or two basic approaches. The first approach, the "dense mode" approach, is based on the assumption that multicast group members are densely distributed throughout the network and that bandwidth is plentiful. The second approach, the "sparse mode" approach, basically assumes that multicast group members are sparsely distributed throughout the network and bandwidth is not necessarily widely available.

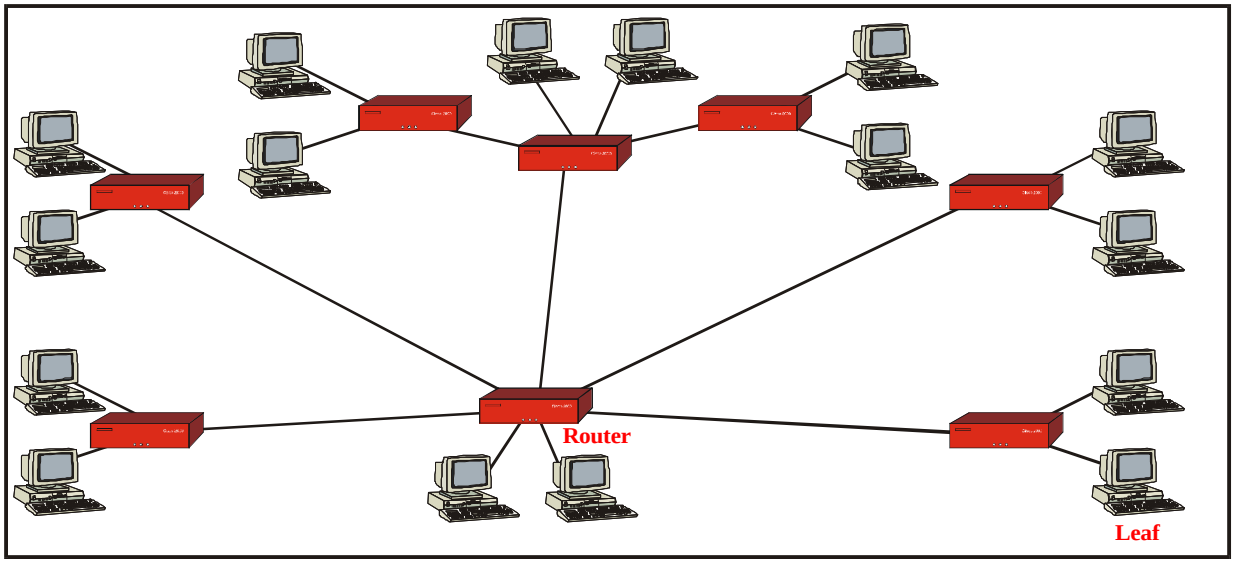

Figure 2-4. Multicast Spanning Tree

Common to both approaches is the need to set up a state in intermediate routers for multicast forwarding. The approaches differ mainly on who initiates the state creation) the sender, the receiver, or the routers themselves. Each approach has its advantages and disadvantages. A significant advantage of the sender-based approach is policy-based routing) the source knows about the policies of nodes it sends information to and can choose a route as desired. Another advantage is that each source can use the multicast route generation algorithm and packet forwarding scheme that best suits it, instead of being forced to use whatever is implemented elsewhere in the network. A disadvantage of the sender-based method is the impact of change in group membership on group dynamism. If there is a change in the membership of the group, the database which contains the group-destination mapping must be updated. In comparison, receiver-oriented approaches appear to be able to accommodate group dynamism more naturally.

For instance, with receiver-initiated trees, a receiver wishing to join a group may generate a policy (i.e., access restriction and QoS), route to the core for that group using its link map, and attach itself to the tree.

2.2.1 DENSE MODE APPROACH

"Dense mode" routing protocols rely on a technique called "flooding" to propagate information to all network routers. In flooding, when a router receives a packet that is addressed to a multicast group, it employs a protocol mechanism to determine whether it is the first time it has seen that packet. If it is the first reception of the packet, the packet is forwarded to all interfaces except the one on which it arrived, guaranteeing that the multicast packet reaches all routers in the network. If the router has seen the packet before, the packet is simply discarded. The flooding algorithm is very simple to implement since a router does not have to maintain routing tables, but only needs to keep track of the most recently seen packet. However, since each router is required to maintain a distinct table entry for each recently seen packet, flooding may not use router memory resources efficiently. Another disadvantage of flooding is that it is generally not suitable for Internet-wide applications since it generates a large number of duplicate packets and uses all available paths across the internetwork instead of just a limited number. The principal dense mode multicast routing protocols offered for consideration by the IETF are the Distance Vector Multicast Routing Protocol (DVMRP), Multicast Open Shortest Path First (MOSPF), and Protocol Independent Multicast-Dense Mode (PIM-DM).

2.2.1.1 DVMRP

DVMRP is the first protocol developed to support multicast routing. It was designed to run over both multicast capable LANs, such as Ethernet, and non-multicast capable routers. DVMRP is based on distance-vector routing distributions and hop-by-hop forwarding. It uses a distance-vector protocol to maintain a current image of the network topology. Distance vector protocols are based on some type of simple measurement (i.e., metric) assigned to each destination in the table, usually the number of hops from the local host. DVMRP assumes that every host on the network belongs to the multicast group. It constructs a different distribution tree for each source and its destination host group. Each distribution tree is the minimum spanning tree from the multicast source at the root to all the multicast receivers as leaves. The distribution tree provides a shortest path between the source and each multicast receiver in the group based on the number of hops in the path. Each router informs its neighbor about its routing table. To establish the network path, the receiving router chooses the neighbor that advertises the lowest cost. The receiving router then adds the path to the low-cost neighbor into its routing table for re-advertisement. As shown in Figure 2-5, the router that has been selected to handle routing for all hosts on its subnet, begins by transmitting a multicast message to all adjacent routers using reverse path multiplexing (RPM). RPM is a multicasting technique in which a router forwards a multicast datagram out on all but the receiving interface, if the receiving interface is one used to

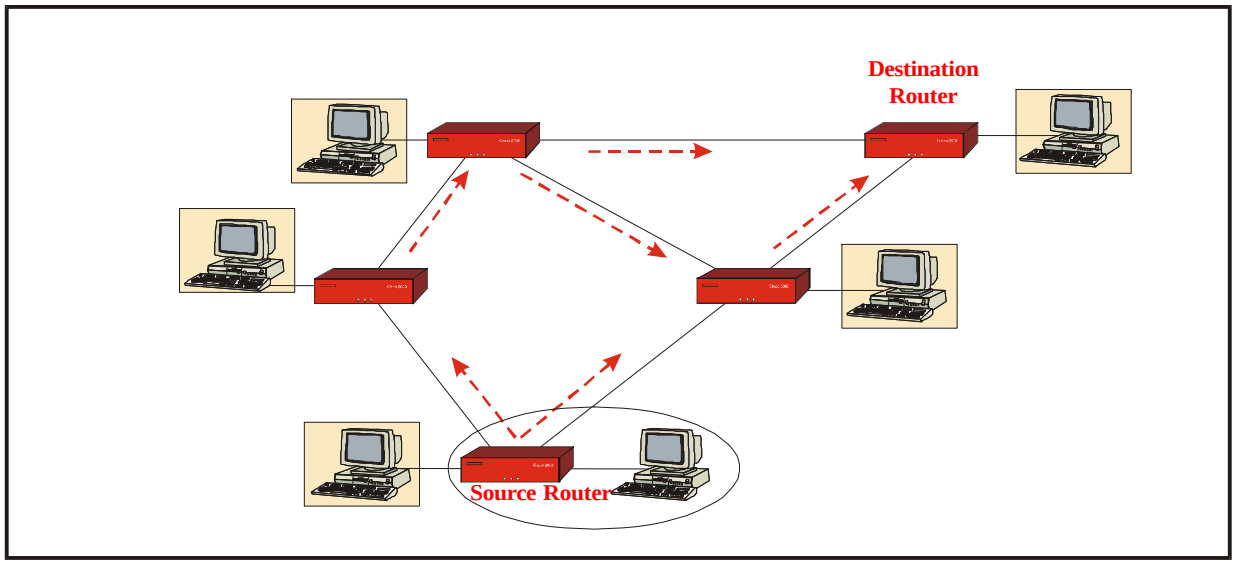

Figure 2-5. DVMRP Routing

forward unicast datagrams to the source of the multicast datagram. The multicast message passes over all router interfaces as it traverses the network. Each adjacent router selectively forwards the message to downstream routers until the message is eventually passed to all multicast group members. The means by which selective forwarding is accomplished is as follows: when a router receives a multicast message, it checks its unicast routing table to determine the interface that provides the shortest path back to the source. If that was the interface over which the message arrived, the router enters appropriate state information in its internal tables to identify the multicast group and to specify interfaces over which messages to that group should be forwarded. The router then forwards the multicast message to all routers other than the one that sent the message.

2.2.1.2 MOSPF

MOSPF is an extension of the Open Shortest Path First (OSPF) link-state protocol. A unicast protocol, it uses the OSPF link-state metric to determine the least-cost path and calculate a spanning tree for routing multicast traffic. The packet forwarding mechanism is hop-by-hop. Network performance parameters that can influence the assignment of cost to a path include the number of hops in the path, requirement for load-balancing, and an application's desired QoS. MOSPF is intended for use within a single routing domain, (e.g., a network controlled by a single organization). With MOSPF, every router has complete topology information on the network and is able to compute the shortest path from any source to any group. MOSPF uses Dijkstra's algorithm) an algorithm used to compute minimum distances from a "source" node to all other nodes in a directed graph) for its "shortest path" computations. If the router doing the computation falls within the tree computed, it can determine which links it must forward copies

to. With the OSPF link-state routing protocol, each router periodically collects information about multicast group membership using the Internet Group Management Protocol (IGMP). The IGMP is used by IP hosts to report their host group memberships to any immediately neighboring multicast routers. The membership and link-state information collected is flooded to all other routers in the routing domain. Routers update their internal link-state information based on information they receive from adjacent routers to maintain topology information of the entire network. If a network link is activated or taken out of service, the MOSPF protocol floods a notification of the change of state throughout the network. All routers note the change and recompute the routes in their routing table. Link-state routing is more reliable than distance vector routing. However, it is also more complex and memory intensive. Group membership information is sent throughout the network, including links that are not in the direct path to the multicast destinations. Thus, like DVMRP, MOSPF appears to be most suitable for small internetworks as an intra-domain routing mechanism.

2.2.1.3 PIM-DM

The PIM routing protocol is presently under development by the IETF. However, PIM support is currently available in some router products. PIM provides a standard multicast routing protocol that supports scalable interdomain multicast routing across the Internet. It is similar to DVMRP in some respects. The PIM architecture maintains the traditional IP multicast service model of receiver-initiated membership, and both protocols employ RPM to construct source-rooted distribution trees. The primary difference between the two is that with PIM, the interdomain multicast routing protocol is completely independent of the unicast routing protocol that is used on the network, while DVMRP relies on specific mechanisms of the associated unicast routing protocol. PIM-DM is one of two PIM operational modes) dense mode and sparse mode. Using two modes of operation provides improved performance both when the group membership in an internetwork is sparse and when it is dense. However, PIM is a complex protocol. A significant limitation of PIM is that the shortest paths are based on the reverse metrics and therefore truly "shortest" only when the links are symmetric. PIM-DM is data-driven and resembles typical multicast routing protocols. However, since PIM-DM is independent of the accompanying unicast routing protocol, data packets which arrive from a router over the proper receiving interface are forwarded on all downstream interfaces until unnecessary branches of the tree are explicitly pruned.

With PIM-DM it is assumed that receivers are densely populated and that the downstream networks want to receive the datagrams forwarded to them. A significant aspect of PIM is that using a "rendezvous point") a router specified to track membership in multicast groups and to forward messages to known multicast group addresses. The rendezvous point allows for the simultaneous existence of shared and source-specific multicast trees. In the steady state, data can be delivered over the reverse shortest path from the sender to the receiver for improved end-to-end delay; the shared tree, which is intended for low-cost multicasting; and the source-based tree intended for low-delay multicasting. When a host wishes to leave a multicast group, its

designated router sends a prune message toward the source (for source-based trees) or toward the rendezvous point (for shared trees). Packets are forwarded on all outgoing interfaces until pruning and truncation occurs. A key consideration in comparing DVMRP and PIM-DM protocols is that DVMRP is more selective when forwarding messages because of the specificity of available topology provided by its own unicast routing protocol. However, PIM-DM is less complex than DVMRP and, although it is likely to be more costly in terms of additional overhead due to some packet duplication, has considerable independence from the unicast routing protocol used on the network.

2.2.2 SPARSE MODE APPROACH

"Sparse mode" does not necessarily imply that the group has few members, just that they are widely dispersed. Since under such conditions, flooding would unnecessarily waste network bandwidth and could potentially cause performance problems, sparse mode routing protocols rely on more selective techniques to establish and maintain trees. The principal sparse mode multicast routing protocols are the Core-based Trees (CBT) and PIM-Sparse Mode (PIM-SM) protocols.

2.2.2.1 CBT

The CBT routing protocol is characterized by one multicast delivery tree with a core router, shared by all senders and receivers of the group. Additional routers may be added for robustness. With CBT, routers send the multicast traffic over the same delivery tree regardless of the source. Routers can join the tree by sending a join message to the core. The chief distinguishing characteristic of CBT is that it is receiver initiated, i.e., receivers wishing to join a multicast group find the tree (or its core) and attach themselves to it, without any participation from the sources. The primary advantages of the shared tree approach is that it typically offers more favorable scaling characteristics than do the other multicast algorithms, and is relatively simple compared to most other multicast routing protocols. However, since traffic from all sources traverses the same set of links as it approaches the core, as traffic increases, CBT could potentially result in delay problems for real-time applications caused by traffic concentration and bottlenecks near core routers. Other potential considerations appear to be that (1) the general performance of the CBT network depends on judicious placement of the cores and the coordination between them, and (2) packets may not traverse the shortest path from the source to their destinations. Figure 2-6 is a nominal illustration of a core-based tree.

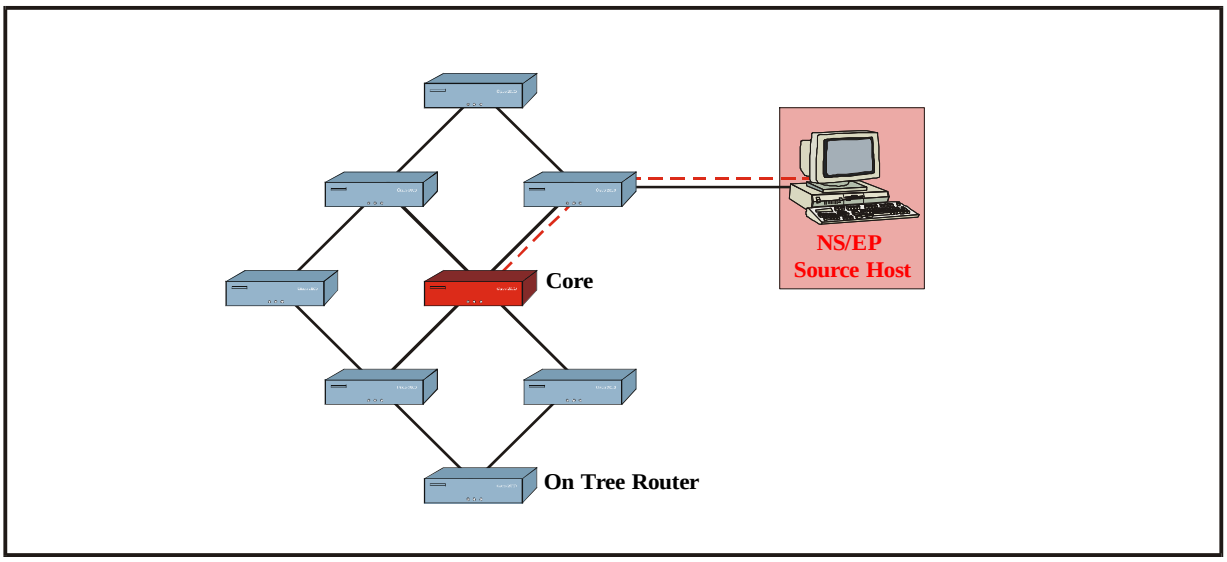

Figure 2-6. Nominal Core-Based Tree

2.2.2.2 PIM-SM

PIM-SM is a protocol optimized for environments where group members are distributed across many regions of the Internet. To receive multicast traffic addressed to the group, routers with directly attached or downstream members join a sparse-mode distribution tree by transmitting explicit join messages. PIM-SM avoids potential scaling issues by limiting multicast traffic so that only routers interested in receiving traffic for a particular group see it.

2.2.3 MULTICAST BACKBONE INTERIM ROUTING APPROACH

Relatively few routers in the Internet currently support IP multicasting. However, as old routers reach the end of their programed lives, most of the new replacement routers do tend to be multicast capable. Several years ago, researchers seeking a way to enable the deployment of a resource and testbed for testing multicasting protocols and applications developed the Internet Multicast Backbone (MBone) as an interim solution. The MBone is an experimental, cooperative volunteer effort spanning several continents. It is a virtual network layered on top of the Internet created by an interconnected set of subnetworks and routers. The MBone supports routing multicast packets without disturbing or altering other Internet traffic. It originated from experiments conducted during IETF meetings in which live audio and video were transmitted around the world, and has been in existence since early 1992. MBone uses a network of routers called "MRouters" that can support IP multicast, augmented with "tunnels" to forward multicast packets between islands of MBone subnets. Tunneling is the practice of encapsulating a message from one protocol in another protocol and using the second protocol to traverse network hops, or

connect islands of multicast routes that are separated by links that do not support IP multicast. Most routers used on the Internet today are unicast（ point-to-point） routers. MBone multicast traffic bypasses the unicast routers on the Internet by the use of software that encapsulates the multicast packets in traditional unicast packets so that unicast routers can handle the information. At the destination, the encapsulation is stripped off and the original message is reintroduced to the network at its destination. As shown in Figure 2-7, tunneling allows multicast traffic to pass seamlessly between two multicast routers connected to unicast routers.

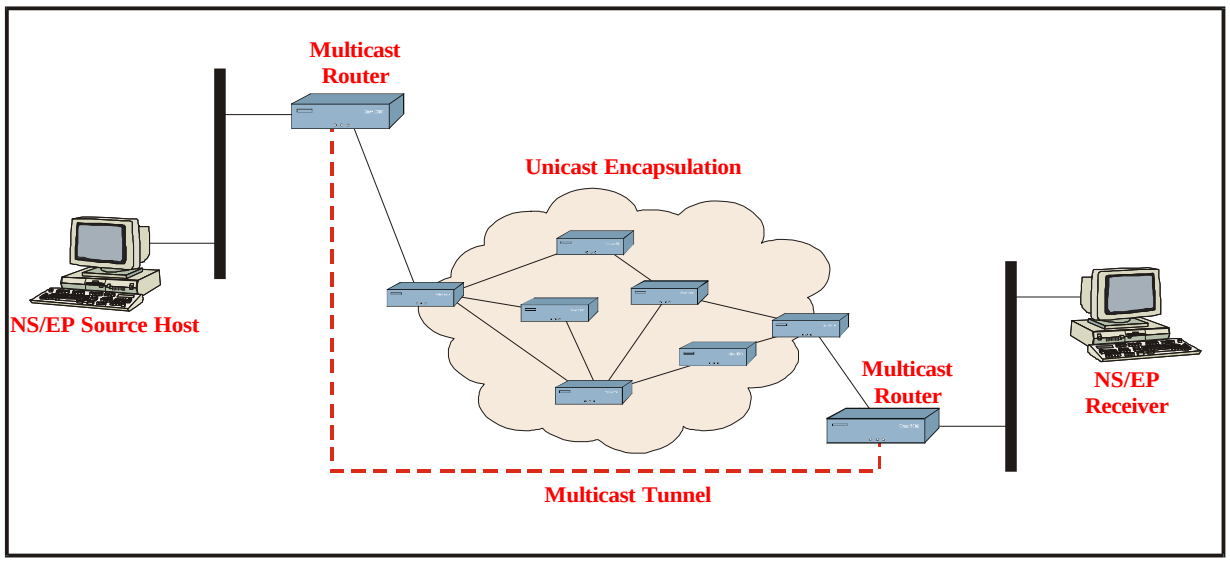

Figure 2.7 IP Multicast Tunneling

2.3 IP MULTICAST DELIVERY

The TCP/IP protocols used on the Internet were designed primarily for the reliable transmission of unicast data with minimal or no delay constraints. TCP works well in this context. However, the Internet, like other packet networks, sometimes loses and reorders packets and delays them by variable amounts of time. This fact, and the inability of the Internet to provide end-to-end QoS guarantees, generally make the current Internet unreliable for the transmission of multimedia traffic. Multimedia traffic is expected to comprise a significant portion of potential multicast traffic. Multimedia traffic exhibits different response characteristics with respect to delay in the delivery of datagrams, and has considerably more stringent QoS requirements. Because of these requirements, TCP is generally not adequate to support real-time multimedia multicast applications.

Network infrastructure devices like routers must support a routing protocol that forwards multicast packets to group members. Multicast applications run on top of the User Data Protocol (UDP) or interface directly to IP via sockets and provide their own customized transport layer. UDP provides for the exchange of datagrams without acknowledgments or guaranteed delivery.

It provides only the minimal transport services of error detection and port multiplexing. Errors or packet loss due to congestion are not recoverable. Consequently, additional protocols are required to provide the necessary transport services. The Internet community (IETF working groups and industry vendors) is working to develop reliable multicast protocols to overcome the limitations cited above. These protocols are currently at various levels of maturity. The principal protocols currently under development include the Real-Time Transport Protocol (RTP), Real-Time Control Protocol (RTCP), ReSerVation Protocol (RSVP), and Real-Time Streaming Protocol (RTSP).

2.3.1 RTP

RTP (a standard that has not yet been formalized) provides end-to-end network transport functions suitable for applications that transmit real-time data over multicast or unicast network services. Such applications can include audio, video, or simulation data. RTP services include payload type identification, sequence numbering, time stamping, and delivery monitoring. Delivery monitoring is by means of a closely integrated control protocol (RTCP. Applications typically run RTP on top of the UDP to make use of its multiplexing and checksum services. Both protocols contribute parts of the transport protocol functionality. RTP supports data transfer to multiple destinations using multicast distribution if provided by the underlying network. While RTP is primarily designed to satisfy the needs of multi-participant multimedia conferences, it is not limited to a particular application. It is designed to allow an application to automatically scale session sizes from a few participants to thousands of participants. Control and measurement applications like storage of continuous data, interactive distributed simulation, and active badging, may also find RTP applicable. RTP is intended to be malleable and is often integrated into the application processing rather than being implemented as a separate layer. Unlike conventional protocols, which accommodate additional functions by making the protocol more general or adding an option mechanism that requires parsing, RTP can be tailored through modifications and/or additions to the headers as needed.

To cope with packet loss, reordering, or delay, the RTP header provides the timing information necessary to synchronize and display audio and video data. It also determines whether packets have been lost or have arrived out of order. In addition, the header specifies the payload type, thus allowing multiple data and compression types. RTP is tailored to a specific application via auxiliary profile and payload format specifications. To set up an RTP session, the application defines a particular pair of destination transport addresses (one network address plus a pair of ports for RTP and RTCP. In a multimedia session, each medium is carried in a separate RTP session, with its own RTCP packet reporting the reception quality for that session. For example audio and video may travel on separate RTP sessions, enabling a recipient to select whether or not he or she chooses to receive a particular medium. RTP does not provide any mechanism to ensure timely delivery or provide other QoS guarantees. It relies on lower-layer services to do this. Nor does it guarantee delivery, prevent out-of-order delivery, or assume that the underlying

network is reliable. For applications requiring such guarantees, RTP must be accompanied by other mechanisms to support resource reservation and to provide reliable service.

2.3.2 RTCP

RTCP works in collaboration with RTP to provide periodic transmissions of control packets to all participants in a session, using the same distribution mechanism as the data packets. Feedback information to the application can be used to control performance and for diagnostic purposes. Request For Comment (RFC) 1889 describes the following four functions performed by the RTCP:

- The primary function of RTCP is to provide information to the application regarding the quality of data distribution. This is an integral part of the RTP's role as a transport protocol and is related to the flow and congestion control functions of other transport protocols. Feedback may be directly useful for control of adaptive encodings as well as to diagnose faults at the receivers.

- The second function of RTCP is to identify the RTP source. RTCP carries a transport-level identifier for an RTP source, called the canonical name or (CNAME). The CNAME is used by receivers to associate multiple data streams from a given participant in a set of related RTP sessions.

- The third function of RTCP is to control RTCP transmission intervals and limit traffic to prevent control traffic from overburdening network resources, and to allow RTP to scale up to a large number of session participants. Since each participant sends control packets to everyone else, each is able to monitor the total number of participants and calculate the rate at which to send packets.

- A fourth optional function of RTCP is to provide a convenient method of conveying minimal amounts of information to all session participants, e.g., a personal name to identify a participant on the user's display. This function may be useful in loosely controlled sessions where participants informally enter and leave the session.

2.3.3 RSVP

RSVP is a protocol developed by the IETF to assist in providing QoS characteristics to communications over an IP network. The name refers to the fact that it allows end-stations to reserve bandwidth on the network and supports requests for a specific QoS from the network for particular data streams or flows. When a host uses RSVP to request a specific QoS from the network, on behalf of an application data stream, RSVP carries the request through the network,

visiting each node the network uses to carry the stream. At each node, RSVP attempts to make a resource reservation for the stream by using the RSVP daemon to communicate with two local decision modules) admission control and policy control. Admission control determines whether the node has sufficient available resources to supply the requested QoS. Policy control determines whether the user has administrative permission to make the reservation. If either check fails, the RSVP program returns an error notification to the application process that originated the request. If both checks succeed, the RSVP daemon sets parameters in a packet classifier and packet scheduler to obtain the desired QoS. The packet classifier determines the QoS class for each packet and the scheduler orders packet transmission to achieve the promised QoS for each stream.

A primary feature of RSVP is its scalability. RSVP scales to very large multicast groups because it uses receiver-oriented reservation requests that merge as they progress up the multicast tree. The reservation for a single receiver does not need to travel to the source of a multicast tree, rather it travels only until it reaches a reserved branch of the tree. While the RSVP protocol is designed specifically for multicast applications, it may also make unicast reservations. RSVP is also designed to utilize the robustness of current Internet routing algorithms. RSVP does not perform its own routing, instead it uses underlying routing protocols to determine where it should carry reservation requests. As routing paths change to adapt to topology changes, RSVP adapts its reservation to the new paths wherever reservations are in place. This modularity does not rule out RSVP from using other routing services. Current research within the RSVP project is focusing on designing RSVP to use routing services that provide alternate paths and fixed paths. RSVP runs over both IPv4 and IPv6. Additional RSVP's features are that it provides opaque transport of traffic control and policy control messages, and provides transparent operation through non-supporting regions.

2.3.4 RTSP

RTSP is an application-level protocol that controls delivery of data with real-time properties. RTSP provides an extensible framework that enables the control of on-demand delivery of real-time data, such as audio and video. Sources of data can include both live data feeds and stored clips. This protocol is intended to control multiple data delivery sessions, and provide a means for choosing delivery channels such as UDP, multicast UDP, and TCP. It also provides a means for choosing delivery mechanisms based upon RTP. RTSP establishes and controls a single or several time-synchronized streams of continuous media. Although interleaving of the continuous media stream with the control stream is possible, it does not typically deliver continuous streams itself. In other words, RTSP acts as a "network remote control" for multimedia servers. The set of streams to be controlled is defined by a presentation description. There is no formal notion of an RTSP connection; instead, a server maintains a session labeled by an identifier. An RTSP session is in no way tied to a transport-level connection such as a TCP connection. During an RTSP session, an RTSP client may open and close several reliable transport connections to the server to issue RTSP requests. Alternatively, it may use a

connectionless transport protocol such as UDP to issue requests. The streams controlled by RTSP may use RTP, but the operation of RTSP does not depend on the transport mechanism used to carry continuous media.

2.4 ASYNCHRONOUS TRANSFER MODE

ATM is a cell-based transfer mode that requires variable length user information of multiple types (e.g., voice, video, or data) to be segmented and reassembled to and from short, fixed length cells. Fixed length 53-byte cells allow cell processing to occur in hardware, thereby reducing transit delay. The first 5 bytes of the cell contain cell-header information, and the remaining 48 bytes contain the payload, or user information. ATM networks are fundamentally connection oriented. That is, they require that a virtual circuit/channel (VC) be setup across the ATM network prior to the transfer of any data. A VC is a communications channel that provides for the unidirectional transport of ATM cells. ATM circuits are of two types: virtual paths (VPs), identified by VP identifiers (VPIs); and virtual channels, identified by the combination of a VPI and a VC identifier (VCI). VPIs and VCIs are unique numerical tags defined by fields in the ATM cell header. A VP is a bundle of VCs, all of which are switched transparently across the ATM network on the basis of the common VPI. However, all VCIs and VPIs have only local significance across a particular link, and are remapped, as appropriate, at each switch.

The basic operation of an ATM switch is to receive a cell across a link from an ATM endpoint (e.g., workstation, routers, digital service units [DSUs], and LAN switches), or another ATM switch on a known VCI or VPI value; look up the connection value and the new VPI/VCI value in a local translation table to determine the outgoing port (or ports) of the connection and the new VPI/VCI value of the connection on that link; and then retransmit the cell on that outgoing link with the appropriate connection identifiers. The local translation tables used by the switches are set up by external mechanisms prior to the transmittal of any data. The manner in which the tables are set up determines the two basic types of ATM connections: permanent virtual connections/circuits (PVCs) and switched virtual connections/circuits (SVCs). PVCs are connections set up by some external mechanism (typically network management) in which a set of switches between an ATM source and destination ATM system are programed with the appropriate VPI/VCI values. ATM signaling can facilitate the set up of PVCs; however, generally PVCs always require some manual configuration. SVCs are connections set up automatically through a signaling protocol. SVCs do not require the manual intervention needed to set up PVCs. All higher layer protocols operating over ATM primarily use SVCs. In subsequent discussions of ATM connections in this report, the principal focus will be on IP multicast over ATM SVCs.

2.4.1 ATM MULTILAYER ARCHITECTURE

The ATM multilayer architecture uses a logical model to describe the functionality it supports. ATM functionality corresponds to the physical layer and part of the data link layer of the OSI reference model. As shown in Figure 2-8, the ATM architecture consists of three functional layers) the physical layer, the ATM layer, and the AAL. The physical layer is the lowest layer. The physical layer is responsible for the transmission of cells between two ATM hosts across a specific physical link. The ATM layer is the next layer above the physical layer. It is independent of the physical layer and the AAL above it. The ATM layer is responsible for a number of functions concerning the cell header, including cell header generation and extraction. Another function performed by the ATM layer is cell multiplexing and demultiplexing. In the transmit direction, cells from individual VPs or VCs are combined into a noncontinuous cell flow which is then passed to the physical layer for transmission. This multiplexing function enables the integration of cell flows from individual connections to be multiplexed over a single physical link. In the received direction, the noncontinuous cell flow is demultiplexed into individual virtual paths or connections based on the VPI or VCI in the cell header. Another very important function performed by the ATM layer is traffic and congestion control. The last layer, the AAL, is the fundamental layer for encapsulation operations. AAL sits on top of the ATM layer. Its primary purpose is to adapt the flow of information received from a higher layer application like voice or data from IP or other upper layers to the ATM layer.

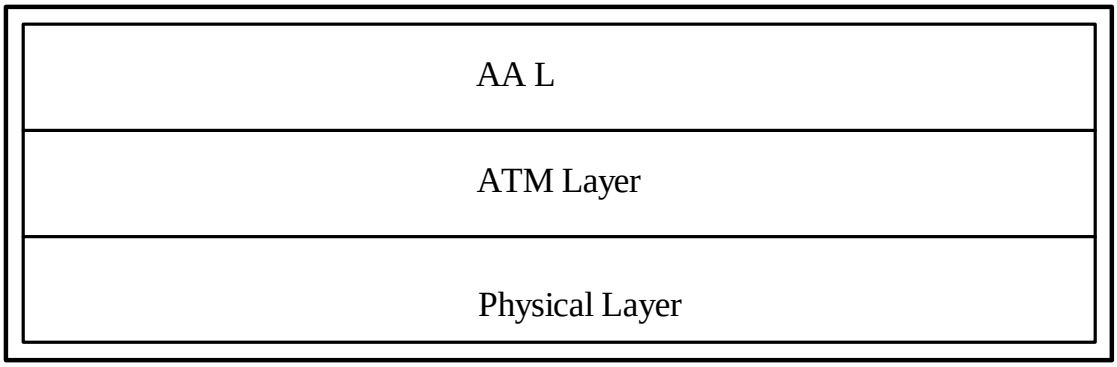

Figure 2-8. ATM Functional Layering

2.4.2 ATM ADAPTATION LAYER

Because ATM was envisioned as a service to integrate many different kinds of applications, four AALs have been defined by the ATM Forum to support different service requirements. In defining the service classes, the criteria used are timing relationship between source and destination, bit rate, and connection mode. AAL-1 typically supports applications that are delay and timing sensitive and require a constant bit rate (CBR). Such applications include uncompressed voice and real-time video. AAL-2 is intended for compressed voice and video in

packetized isochronous format. Compression allows for a variable bit rate (VBR) without losing audio and voice quality. However, timing is still important. AAL-2 is still under study within international standards bodies. AAL-3/4 was originally two separate AALs intended to support transport of VBR, delay-tolerant, connection-oriented and connectionless data traffic requiring some sequencing and error detection support. However, as the specifications evolved, it became apparent that both services required similar procedures and, as a result, the specifications were merged to become the AAL-3/4 standard. AAL-3/4 supports both connection-oriented and connectionless data transport, and two modes of operation) message mode and streaming mode operations.

Message mode service is used to transfer one frame of information from a higher layer application. Streaming mode service is used to transfer one or more frames of information from a higher layer application. With the streaming mode, the frames of information may be separated in time. AAL-3/4 supports both assured and nonassured delivery. AAL-5 was designed specifically to support transport of VBR, delay-tolerant, connection-oriented data traffic requiring minimal sequencing and error detection support. This is typically the type of data found in current LANs. It evolved because AAL-3/4 was considered too complex and inefficient for LAN traffic. Like AAL-3/4, AAL-5 supports both the message and streaming mode operations as well as assured and nonassured delivery options. Its connection-oriented mode guarantees delivery of data by the servicing applications and does not add any cell overhead.

SECTION 3.0

ISSUES: IP MULTICASTING OVER ATM NETWORKS

The IETF IPATM Working Group Charter states its intention to "...focus on the issues involved in running internetworking protocols over ATM networks." According to the charter, "The final goal of the Working Group is to produce standards for the TCP/IP protocol suite and recommendations which could be used by other internetworking protocol standards...". [1] There are significant differences in the way IP and ATM operate. The principal difference involves the nature of the two technologies, particularly the methods used for handling the transport layer services of connection management, data transfer, and flow control. Connection management includes establishing and terminating connections between transport users, identifying each connection, and negotiating values of all needed parameters. Data transfer involves the reliable delivery of transparent, in-sequence data between users without duplication or missing elements. Flow control involves traffic control measures taken by the network to manage congestion in order to assure that user traffic does not saturate the network or exceed network capacity.

IP provides a connectionless service which allows the transfer of information divided into packets, or datagrams, among service subscribers without the need for end-to-end establishment of a transmission path. Each datagram is transmitted individually and can even follow different routes to its destination. Once all the datagrams forming a message arrive at the destination, they are recompiled into the original message. ATM is a connection-oriented transport service which requires that a VC be established between the sender and receiver before traffic organized into fixed-sized cells can be transmitted. Once a VC is established, the ATM protocol selects a physical route from source to destination and enters the information in a route table in the switch. All switches along the pathway make table entries so they can route packets accordingly. All messages for that connection follow the same path to the destination. Use of the services of ATM for IP multicasting requires a mechanism which makes the difference in characteristics between ATM and IP transparent to the ATM network. Figure 3-1 shows a nominal IP-ATM internetwork with IP packets routed from a source over an "ATM cloud" to a multicast router and onward to receiving hosts. Achieving the transparency objective cited above involves significant issues of address registration, address resolution, encapsulation, connection establishment, and routing.

3.1 ADDRESS REGISTRATION

Each device within a network must have a unique and identifiable address in order to receive and transmit messages. To establish an ATM connection at the user-network interface (UNI), both the user and the network must know the ATM addresses in effect at the UNI. These addresses are used in the Calling Party Number information elements of signaling messages sent by the

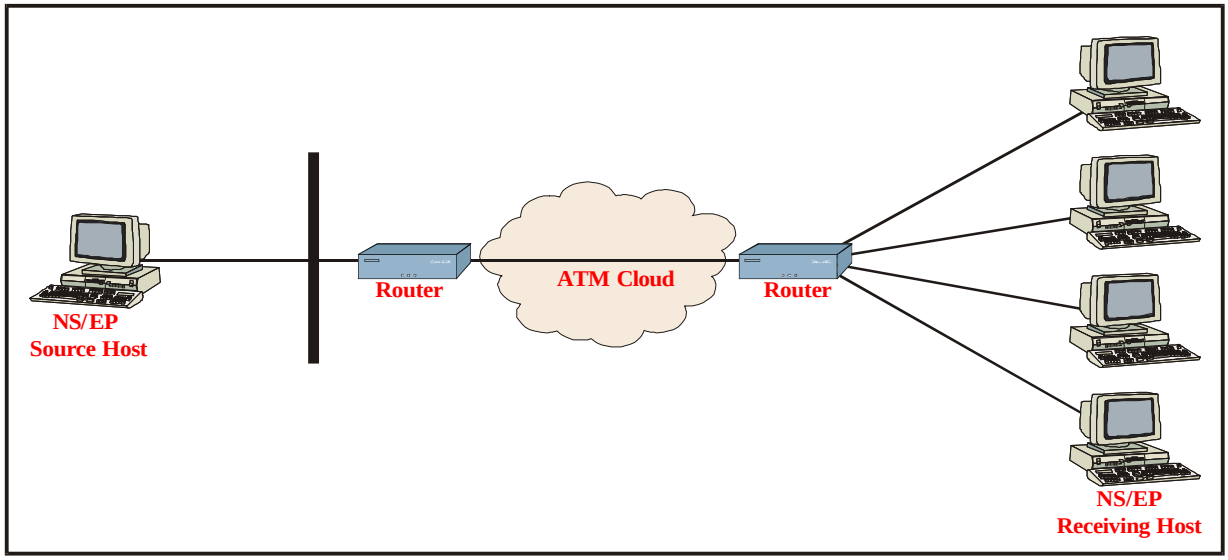

Figure 3-1. IP Multicast Over ATM

user, and in the Called Party Number information elements of signaling messages sent to the users. Address registration is the dynamic exchange of network routing prefixes on the network side and end system identifiers on the host side. Existing network layer protocols (e.g., IP) have their own addressing schemes and associated routing protocols. However, in an integrated IP over ATM environment, the ATM layer is decoupled from IP and is defined by its own addressing structure. The ATM addressing space is logically disjointed from the addressing space of whatever protocol runs over it and typically would not bear any relationship to it. Hence, all protocols operating over ATM require some form of address registration and the use of an ATM address resolution protocol to map higher layer addresses to their corresponding ATM addresses. To facilitate the administration and configuration of addressing in an integrated IP/ATM environment, the IETF Networking Working Group and the ATM Forum Technical Committee are currently studying address registration and address resolution issues.

IP uses a 32-bit address to identify a host computer and the network to which it is attached. IP addresses are classified by their format. Four format classes are permitted) classes A, B, C, or D. The first bits of the address specify the format of the remainder of the address field in relation to the network and host fields. Class A addresses are used for networks that have a large number of hosts) a maximum of 2^{24} hosts. Class B addresses are used for networks of intermediate size. Class C networks contain fewer than 256 hosts (2^{8}). Class D addresses are reserved for multicasting. For ease in reading, the IP address is depicted in the dotted decimal form of X.X.X.X, where each X represents an eight-bit byte. An example IP address is 171.16.17.55. The Internet Information Center (InterNIC) is the combined name for the providers of IP address registration, information, and database services to the Internet.

The ATM Forum has defined three types of ATM End System Addresses (AESAs) for use in private networks. An AESA is a 20-byte number that is used to identify an ATM endpoint from any other endpoint in a global network. The AESA structure is derived from the address structure of a generic OSI network addressing standard) the Network Service Access Point (NSAP). For this reason, AESAs are often referred to as ATM NSAPs. The three ATM NSAP AESA formats are the Data Country Code (DCC) AESA, the International Code Designator (ICD) AESA, and the E.164 AESA. All NSAP format ATM addresses consist of three components: an Authority and Format Identifier (AFI), which identifies the type and format of the Initial Domain Identifier (IDI); the IDI, which identifies the address allocation and administration authority; and the Domain Specific Part (DSP), which contains actual routing information. The Federal Government has selected the ICD AESA addressing format for use in its Government Open Systems Interconnection Profile (GOSIP) protocol procurement specification.

The major difference in the various types of ATM NSAP AESA addresses is in the authority that assigns them to users. Assignment of AESA addresses is accomplished by a series of authorities. A top level authority assigns a portion of the prefix and delegates to a lower level authority to assign the remainder of the address. With ATM, there are several registration authorities from which network managers can acquire unique network name and address space. In the United States, ATM code points may be obtained from the National Institute of Standards and Technology (NIST) and the American National Standards Institute (ANSI). Basically, NIST or ANSI assigns a 3-octet organization identifier code point. This field follows 4 octets of identification bits which identify the country and registration authority (e.g., ANSI). The owning organization is then responsible for the encoding of the remaining 6 octets in the ATM network part of the ATM address.

In ATM networks, both individual and group addresses are used to identify endpoints. An individual address is used to identify a single ATM end system, whereas an ATM group address is used to identify one or more ATM end systems. An ATM end system may join or leave a group at any time by using the client registration and deregistration procedures outlined in the ATM Forum's *Interim Link Management Interface Specification, Version 4.0*. End nodes maintain the lookup tables that translate addresses into circuit paths. These circuit path lookup tables differ at every node and are maintained in a "quasi"real-time fashion by a routing protocol. Each node on the network must register its address in the address register in the router. The router maintains a regularly updated register of addresses and tables for all nodes in the network. VP routing involves the translation of VPI values of the incoming VP links into the VPI values of the outgoing VP links. A VP is assigned a specific value of VPI each time a VP is switched in the network.

The ATM Forum has specified E.164) the International Telecommunication Union-Telecommunications Standardization Sector (ITU-T) standard that specifies the telephone number-type format used for the Integrated Services Digital Network (ISDN)) as the addressing standard to be used in ATM public networks. This election allows legacy public

telecommunications operator (PTO) networks to be migrated to ATM without having to undergo major re-numbering. The E.164 address format is the same as the format used in the public telephone network. E.164 addresses identify interfaces, not endpoints as in the case of ATM NSAP AESA addresses. The address prefixes within E.164 are assigned on a world zone, and then a country-by-country basis. Each country then determines its own numbering plan. The E.164 format limits the maximum length of an address to 15 digits) the first one to three of which are assigned as a region code to designate a specific country. Within a given region, a Numbering Plan Authority decides how the remaining digits after the region code are to be structured and assigned. In the United States, the well-known North American Numbering Plan (NANP), with its three-digit numbering plan code or area code, three-digit central office code or prefix, and four-digit line number make-up, provides an excellent example of an E.164 address structure. ISO recommendations describe a hierarchical structure for the NSAP address.

3.2 ADDRESS RESOLUTION

Address resolution is used to resolve differences between different addressing schemes. Typically, address resolution specifies a technique for mapping network layer (Layer 3) addresses to data link layer (Layer 2) addresses. In a classical IP environment, the source and destination address fields of the IP datagram header contain the addressing information needed to route datagrams in a connectionless network. IP routers using static or dynamic routing tables attempt to match the network address contained in the header of a datagram with a network address contained in the routing table. If the destination node is on its local network, the datagram is forwarded directly to the destination host. If the destination node is on some other network, the datagram is forwarded to the IP local router for forwarding as appropriate. The IP multicast model is a receiver-initiated model. Receivers wishing to subscribe to a multicast group use the IGMP protocol to inform their local router. Routers, using multicast routing protocols, e.g., DVMRP, MOSPF, or PIM, disseminate membership information to all other routers in the routing domain. An IP sender wishing to transmit data to a multicast group sends the IP packets to the IP address of the multicast group. With IP over ATM, the IP addressing information is encapsulated within the ATM cell along with the rest of the datagram and becomes transparent to the ATM network. Consequently, IP multicasting over ATM networks requires a mechanism for mapping IP multicast group addresses to corresponding ATM addresses. The mechanism used to manage the mapping of IP multicast group addresses to corresponding ATM addresses for IP packet forwarding is the Multicast Address Resolution Server (MARS). The MARS acts as a registry, associating IP Layer 3 multicast group addresses to one or more ATM interface addresses representing the group's members. Each ATM-based host and router client communicates with the MARS by using a globally known VC. The MARS may reside within any ATM endpoint that is directly addressable by the endpoints it is serving. When a new host is added to the network, it must register with the MARS to provide a table entry that maps its corresponding IP address to its ATM address. The MARS manages a cluster of ATM-attached endpoints. The IETF defines a cluster as "the set of ATM interfaces choosing to participate in direct ATM connections to achieve multicasting of ATM AAL Service Data Units

(SDUs)) units of interface information whose identity is preserved from one end of a layer connection to the other) between themselves."

Endpoints wishing to join a multicast cluster must be configured with the ATM address of the node on which the cluster's MARS resides. Each IP/ATM interface must keep state information regarding the ATM addresses of each leaf node on each point-to-multipoint connection it has opened. Traffic between interfaces to different clusters passes through an inter-cluster device, e.g., an IP multicast router with logical interfaces into each cluster. The distribution of multicast group membership information between the MARS and the endpoints is accomplished through messages. Endpoint address resolution entities query the MARS when a network level address needs to be resolved, and informs the MARS when they need to join or leave a particular group. It should be noted that an endpoint decision to join or leave a group is a local issue. It has no effect on other members of the multicast group. Currently, the ATM Forum defined Integrated Link Management Interface (ILMI)) an interim specification for network management functions between an end user and a public or private network and between a public network and a private network) is the only standardized way to automatically configure end system addresses. However, the ILMI address registration process only works for UNI interfaces for single end system addresses at a time. The automatic assignment of sub-network group addresses needed to permit point-to-multipoint connections to be setup to multiple leaves in one request is not supported.

A potential concern with the ATM Forum's MPOA model multicast arrangement is that while MARS appears to be an effective approach for small environments, several studies have concluded that it will not scale well to large networks where multicast traffic must transit to a single multicast server. Other concerns are that (1) the current proposal for MARS does not appear to provide sufficient flexibility to handle the range of new application requirements for QoS and traffic behavior, and (2) the requirement to establish dedicated VCs between the root and each recipient leaf node, and the provisioning and management of the MARS mapping information translates into considerable overhead. Discussion is currently ongoing within the IETF concerning ways by which the overhead associated with the establishment and maintenance of separate multicast VCs might be avoided. It is generally agreed that for certain long-duration applications requiring QoS guarantees, the establishment of multicast VCs can be justified. However, for relatively short-duration applications lacking the requirement for QoS guarantees, the use of some type of shared service) such as could be provided by a multicast server (MCS)) is also being discussed as an option. The current focus of the discussion is on the relative merits of using "VC meshes") overlaid point-to-multipoint connections) versus MCSs to support network layer multicasting over ATM. Both seem to offer advantages and disadvantages. IETF Network Working Group RFC 2022, entitled *Support for Multicast over UNI 3.0/3.1 based ATM Networks*, lists the following tradeoffs for each approach:

The VC Mesh: With the multicast VC mesh, each source establishes independent point-to-multipoint VCs to the set of leaf nodes it wishes to send messages. Interfaces for the leaf nodes originate and terminate VCs, as appropriate, for each active leaf node. The term "VC mesh" is

used to describe the resulting crisscross VC pattern. The VC mesh lacks the obvious single congestion point of an MCS. Throughput is likely to be higher, and end-to-end latency lower, because the mesh lacks the intermediate AAL-SDU reassembly that must occur in MCSs. The underlying ATM signaling system also has greater opportunity to ensure optimal branching points at ATM switches along the multicast trees originating on each source. Resource consumption will be higher. Every group member's ATM interface must terminate a VC per sender (consuming on-board memory for state information and requiring and buffering in accordance with the vendor's particular architecture). With an MCS, only two VCs (one out, one in) are required. With a multicast server, the allocation of VC-related resources is also lower within the ATM internetwork.

The Multicast Server: With the MCS, each source establishes a VC to the MCS. The MCS establishes a point-to-multipoint VC to the leaf nodes. AAL-SDUs arriving on incoming VCs are reassembled by the MCS and queued for transmission on a single outgoing point-to-multipoint VC. Since AAL-5, the most common AAL for data, currently does not support cell level interleaving/multiplexing of different AAL-SDUs on a single outgoing VC, reassembly of incoming AAL-SDUs by the MCS is required. Consequently, AAL-5 does not support multicasting. With regard to the signaling load, the MCS has the advantage over the VC mesh when faced with dynamic sets of receivers. When using an MCS, every time the membership of a multicast group changes, i.e., a leaf node needs to be added or dropped, only a single point-to-multipoint VC needs to be modified. This generates a single signaling event across the MCS's UNI. When a membership change occurs in a VC mesh, signaling events occur at the UNI of every traffic source. The transient signaling load is determined by the number of sources. However, MCS introduces a "reflected packet" problem which requires additional AAL-SDU information to be carried in order for network layer sources to detect returns of their own AAL-SDUs.

3.3 ENCAPSULATION

Encapsulation is an operation which allows a transport network to carry traffic using non-similar protocols. Encapsulation encloses data formatted by protocols operating at the upper layers of the OSI layered model, e.g., IP datagrams or protocol data units (PDUs), within another protocol that performs lower layer bearer services, e.g., an ATM cell, in order to transport the encapsulated data across a network for which the original protocol was not designed. Ideally, the transport network does not become involved with either the syntax or the format of the transported traffic. The function of mapping IP PDUs into the information field of the ATM cell and vice versa is performed in the AAL. When a VC is created, a specific AAL type is associated with the VC. The AAL type is known by the VC endpoints via the call setup mechanism and is not carried in the cell header. To invoke encapsulation operations, the sender furnishes the network with a specific identifier to distinguish the type of traffic that is to be transported through the network. This is required so that interworking units (IWUs), i.e., routers, used to perform relaying functions between networks, and receiving user equipment can invoke support

procedures that apply to the specific protocol family. In most systems, services invoked at a layer are requested by the upper layers which pass transactions to the next lower layer to identify the type of service needed. Figure 3-2 provides a simplified illustration of how user traffic can be interpreted and transported through the ATM network. The manner in which the end-user stations communicate with the routers is not defined by ATM, since the information flow between the end-user station and the router is not part of the ATM interface. The interface with the ATM network occurs at the UNI. However, the user station-to-router operation is well understood and defined in existing standards. The router need only map information received from the user stations into the ATM AAL PDU at the originating router and perform a complimentary and reverse operation at the terminating router.

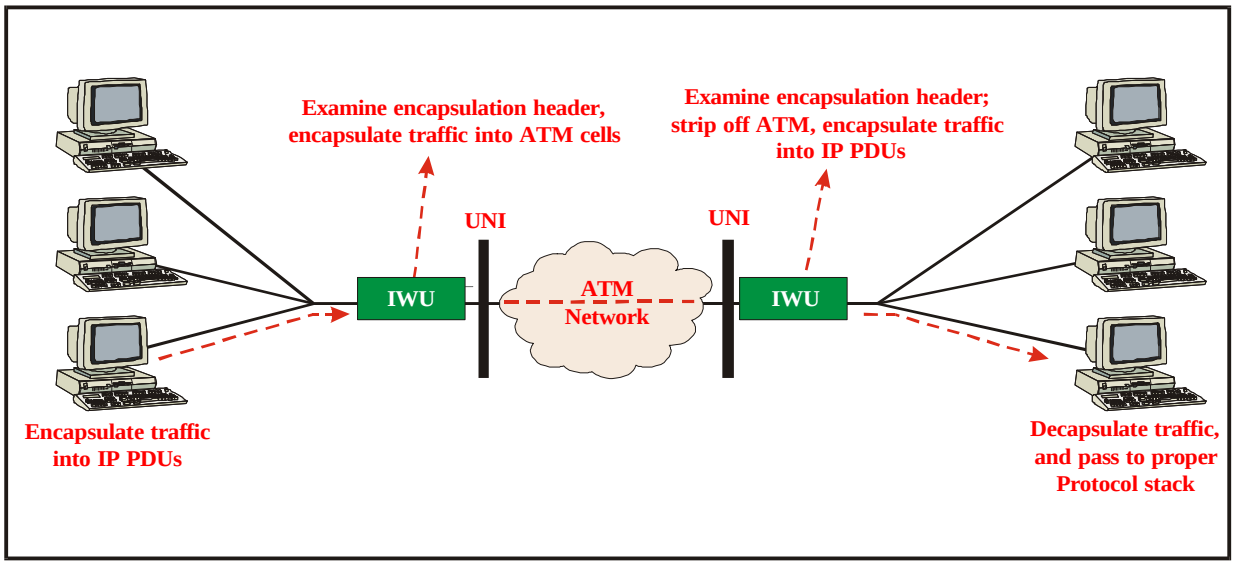

Figure 3-2. ATM Transport Network

Because ATM is a cell-based transfer mode, variable IP multicast data must be formatted into short, fixed length cells prior to transport across the ATM network. IP packets are often considerably larger than ATM cells. Consequently, there is generally a requirement to reduce the size of the IP multicast packets to adapt them to the ATM cell size. As part of the encapsulation function, fragmentation) the partitioning of IP datagrams into parts) is the method used to separate datagrams that are too large for the transport network technology to support. In the transmit direction, the encapsulation function includes the generation of an appropriate cell header for the information field by the ATM layer, less the header error control (HEC) which is the responsibility of the physical layer. The encapsulation function may include translation from a service access point (SAP) identifier to a VPI and VCI. A SAP is a physical interface between the layers of the OSI model through which lower layers provide services to the higher layers passing over the PDUs. In the receive direction, the decapsulation function performed by the ATM layer includes the extraction of the ATM cell header and the passing of the cell information field to the AAL for mapping into appropriate PDUs. The encapsulation function introduces an element of

delay which could have an impact on the response characteristics of multimedia traffic. Multimedia traffic is expected to comprise a significant portion of future potential NS/EP multicast traffic.

With IP multicast over ATM, the IP access protocols terminate at the router gateway. The ATM protocol platform creates an envelope in which the IP protocol formatted data is transported transparent to the network. At the destination endpoint, the data is removed from the envelope and returned to its original form. To allow the receiver to properly process the incoming PDU, the information necessary to identify the protocol of the routed or bridged PDU is carried in the Payload Field of the AAL. RFC 1483, *Multiprotocol Encapsulation over ATM Adaptation Layer 5*, describes two methods of sending connectionless traffic, such as IP, over ATM. The first method supports multiplexing multiple protocols over a single ATM VC. Using this method, Logical Link Control (LLC) encapsulation, the protocol of a carried PDU is identified by prefixing the PDU with an IEEE 802.2 LLC header. A logical link is an abstract representation of the connectivity between two logical nodes. The second method, Based Multiplexing, differs from LLC encapsulation in that the VC is terminated directly at a Layer-3 endpoint. VC-based multiplexing assumes that each protocol is carried over a separate ATM VC. Higher layer protocol multiplexing is done by ATM VCs, requiring that the transported protocol be identified by the ATM VC. AAL endpoints would be Layer-3 protocol entries, requiring that a VC carry one protocol only. In a multiprotocol environment this scheme would use a number of VCs. The advantage of LLC encapsulation is that multiprotocols can share a single VC thus limiting the number of VCs required in IP multicast and multiprotocol environments. It is envisioned that VC-based multiplexing will be dominant in environments where dynamic creation of large numbers of ATM VCs is economical and needed. However, LLC encapsulation may be desirable when it is practical for one reason or another to have a separate VC for each carried protocol.

Cell replication is done within the network by the ATM switches where the connection splits into two or more branches. Such connections are unidirectional, permitting the source end-system, or root, to transmit to the destination end-systems, or leaves, but not to the root or each other on the same connection. The capability of multiple end-systems to receive and transmit data to other multiple systems is common in many shared LAN technologies such as Ethernet or Token Ring. In shared media LAN technologies, all nodes on a single LAN segment must necessarily process all packets sent on that segment. However, this is not the case for ATM networks. AAL5, the most common AAL used to transmit data across ATM networks, does not have any provisions within its cell format for the interleaving of cells from different AAL5 packets on a single connection. If a leaf node transmits an AAL5 packet onto the connection, it is received by both the root node and all other leaf nodes. At these nodes, the packet sent by the leaf could be interleaved with packets sent by the root and possibly other leaf nodes. Consequently, cells sent to a particular destination across a particular connection must be received in sequence, with no interleaving between the cells of different packets on the same connection, or the destination reassembly process would not be able to reconstruct the packets.

3.4 CONNECTION ESTABLISHMENT

Since ATM is connection-oriented, a connection request needs to be routed from the requesting node through the ATM network to the destination node, much as packets are routed within a packet-switched network. In ATM-based networks, there are basically two fundamental types of connections) point-to-point connections and point-to-multipoint connections. Point-to-point connections are connections between two ATM end-systems. Such connections may be either unidirectional or bidirectional. Point-to-multipoint connections connect single source end-systems to multiple destination end-systems. Such connections are unidirectional, permitting the root to transmit to the leaves, but not allowing the leaves to transmit to the root or to each other on the same connection. Connection establishment for SVCs is by mutual agreement and can be setup using a simple set of user commands. ATM switches are preconfigured to receive any signaling packets sent across the connection and pass them to a signaling process associated with the switch. It is the signaling and call control processes that setup the connection through the switches. Signaling is initiated by an end-system. The signaling is routed through the network, from switch to switch, setting up the connection identifiers as it goes. This process provides the switches with information used for routing and making statistical measurements for use when accepting new connections. The routing of the connection request, and the subsequent flow of data, is governed by the ATM routing protocols. The user defines the endpoints when the call is initiated. For unicast, point-to-point, connections, the signaling protocol used is an exchange of messages between the caller and receiver across an adjacent ATM switch. The called and calling party numbers are coded in accordance with ATM Forum address rules. Message information is used to build, maintain, and clear the connection. The messages themselves are segmented into cells at the segmentation and reassembly (SAR) sublayer of the AAL and then transported over standard signaling channels. Figure 3-3 shows the flow of messages during connection establishment.

To establish a connection and send data, a user sends a SETUP message signal that specifies the called party address. The called party can send a close command to return to the idle state, or may open a connection by sending a CONNECT message back to the calling party. If a CONNECT message is received, the calling party ascertains the existence and allocation of VPI/VCI values and returns a connection acknowledgment (CONNECT-ACK) message to the called party's signaling equipment. The connection circuit is coded to indicate what VCI and VPI values have been assigned by the network. The CONNECT message also allows the calling party to enable itself to receive traffic. Since the connection acknowledgment has local significance only on the called party side, the called party does not know that its initial data is received by the calling party until it receives some end-to-end indication (IND) from the calling party. The READY-IND message is sent by the calling party after it has performed its housekeeping functions and is ready to receive frames on the same VC connection (VCC). A VCC is the concatenation of virtual channel links (VCLs) that extends between the points where the ATM service users access the ATM layer. It is at the VCC endpoints that the ATM cell payload is passed to, or received from, the users of the ATM layer for processing.

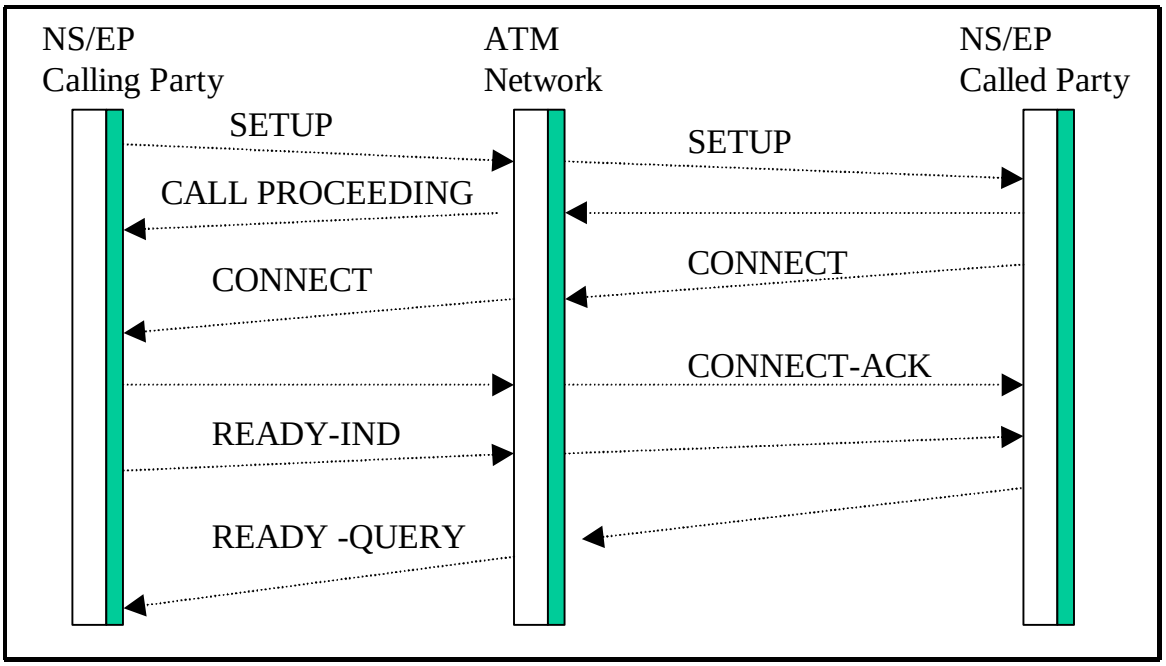

Figure 3-3. Connection Establishment Message Flow

After sending the READY-IND message, the calling party may begin to send traffic immediately on the VCC. Upon receiving the CONNECT-ACK message, the called party starts a timer. The use of the timer overcomes the problem with the local nature of the CONNECT-ACK message. Upon expiration of the timer and if the called party does not receive the READY-IND message, it can send a READY-QUERY message to the calling party on the VCC. These actions are taken to ensure that both parties are aware of the connection. The calling party must respond to the READY-QUERY message with a READY-IND message. The presence of these components is mandatory in every message. A message may contain different information elements (IEs) depending on the type of message. Each message sent across the network includes the following components: protocol discriminator, reference values, message type, and message length. Connection termination like connection establishment is by mutual agreement. Either side or both users of a connection may initiate a connection termination command following the same general procedures. RELEASE and RELEASE COMPLETE messages are used instead of the messages cited above.

Multicast connections are implemented in a slightly different manner than unicast operations. Multicast connections are supported by a collection of sender-initiated point-to-multipoint, unidirectional connections and associated endpoints. This arrangement requires that the router know each intended recipient, and explicitly establish a connection between itself as the root and each recipient as a leaf node. A typical connection setup is achieved by first establishing a point-to-point connection between the root node and one leaf node. After the first set up is complete,

additional leaf nodes are added to the connection by "Add-Party" message requests through the root nodes. Leaf nodes may be added or dropped at any time after establishing the connection. The calling party/host sends the frames to the MARS and not to the client, by using the point-to-point connection cited above. During address resolution, the ATM address of the MARS is provided and not the ATM address of the end users. Information flowing from the source is replicated at the router. Three methods have been proposed for the flow of information onward to the end-user: the VC mesh, the MCS, and VP multicasting.

In the VC mesh method, separate point-to-point circuits connect the MARS to all end-users/clusters of members. All nodes in the multicast group establish a point-to-multipoint connection with each other node in the group, thereby becoming a leaf in the equivalent connections of all other nodes. Hence, all nodes can transmit to, and receive from, all other nodes. Use of the VC mesh requires each node maintain N connections for each group, where N is the total number of transmitting nodes within the group. It also requires a registration process for telling nodes that join a group what the other nodes in the group are, so that they can form there own point-to-multipoint connections.

With the second method, the MCS, the MCS establishes point-to-multipoint connections to the final destination end-user. In turn, end-users wishing to receive multicast traffic, need only connect to the MCS. The MCS method is more scalable than the VC mesh, but requires a centralized resequencer, which can be both a potential bottleneck and a single point of failure. Certain ATM switch architectures use buffered switching networks with dynamic routing, a form of cell routing in which individual cells in a VC are independently routed to the proper output port. Dynamic routing increases the possibility of cells getting out of order on their way through the switch. The use of resequencers in the output port processor restores the proper cell ordering. However, standard resequencing methods introduce delay in the delivery of cells to the output link. This added delay could become an issue for certain applications. The MARS works as a group information holder in a multicast cluster and is responsible for tracking the IP group membership information across all cluster members. It also provides on-demand associations between IP multicast group identifiers and ATM endpoint addresses.

VP multicasting, a third connection method under discussion in certain forums, could provide multipoint-to-multipoint VP connections linking all nodes in a multicast group. Each node would be given a unique VCI value within the VP. Interleaved packets could then be identified by the unique VCI value of the source. However, this mechanism requires a protocol to uniquely allocate VCI values to nodes, and at present no such protocol exists. Additionally, it is also not clear whether current SAR devices could easily support such a mode of operations. It appears that connections via both a VC mesh and an MCS could increase packet delay due to the setup time required by the lengthy connection establishment procedures. Because of the need to traverse bidirectional connections for an MCS versus a unidirectional connection for the mesh, it seems that delay would be longer for an MCS than a VC mesh. Such delay might be intolerable for multimedia applications such as the real-time voice and video on-demand, which is expected to make up a large portion of NS/EP multicast traffic during certain emergency situations.

3.5 ROUTING

As a connection-oriented switching fabric, ATM routes are established at connection setup time and remain in place until the connection is terminated. An ATM cell only carries information identifying the connection and no information about the actual source and destination of the cell. To forward cells, an ATM device consults a list of the established connections that map to the next hop device, without checking the final destination. Network ingress and egress points provide Layer 3 routing functions, while standard ATM routing protocols are used to carry traffic through the network. Routing functions for VPs are performed at the VP switch/cross-connect. Routing involves translation of VPI values of incoming VP links into VPI values of outgoing VP links.

Routing IP multicast packets across an ATM network is a topic of current discussion both in the ATM Forum and IETF networking groups. The two network entities represent two independent routing approaches and hierarchies that make it difficult to adequately coordinate routing across the two topologies. IP multicast routing is unaware and independent of the ATM topology. For a large IP network running over a large ATM network, the user needs to install and manage two independent routing hierarchies (one for ATM, and one for IP). To forward an IP multicast packet to destinations across an ATM network, the packets traverse the network of IP routers following the path specified by the standard routing computation until they reach an IP router at the ATM network interface. At this node, the routing function of the IP subnet is terminated. The ATM network forwards the IP packet across an existing PVC if available, or buffers the packet while setting up an SVC to the associated address, if an SVC is required. The created SVC remains for a specified period of time and closes if no traffic is passed through it. It appears that delay in forwarding packets caused by the SVC process itself could be a significant consideration for some multicast applications.

Because of differences in approaches in specifying QoS routing in both topologies, a potential complication exists in the transport of IP multicast packet over an ATM network. Ideally when an IP host commits to provide a specified type of service for an application, it must be able to request an appropriate QoS from the ATM network using the ATM service model. Most unicast and multicast IP routing protocols compute the shortest path to a destination based solely on a hop count or metric. However, no current IP multicast protocol takes into consideration the wide range of levels of QoS that are available in ATM networks. In many routing protocols computing all the routes for just the shortest path for a large network requires a large number of computations. Consequently, repeating the process for multiple QoS levels in ATM networks might be cost prohibitive. Because of the broad range of options, a potentially complex mapping function must be performed for the IP layer to meet its commitments. For example, traffic in an IP subnet with a reservation request from a host would at some point encounter the edge of the ATM cloud. At this point, either a new connection setup across the ATM cloud is required, or the router must determine if it is possible to carry the requested traffic over an existing VC. If the

ATM cloud cannot create a new connection as requested, the result may be an admission control failure which causes the router to deny the reservation request. Although mechanisms exist today for traffic prioritization on router-based IP networks, they do not appear to be well suited to the demands of ATM.

The IETF is examining several mechanisms by which QoS specifications for IP multicast can be translated into QoS specifications that are meaningful for an ATM network. The principal mechanisms are discussed in Section 2.3. Included are both receiver-initiated mechanisms, and depending on the distribution environment, receiver/sender-initiated mechanisms. However, it appears from an initial examination that receiver-initiated mechanisms may not be suitable for ATM network use. In ATM networks, resource reservations are made at connection setup, using UNI and network-network interface (NNI) signaling protocols. The differences between receiver-initiated mechanisms and ATM state establishment could raise potential problems in that the service priorities established at the IP subnet may not be carried through the ATM network, thereby creating network inefficiencies in executing IP service contracts.

LIST OF REFERENCES

1. Internet Engineering Task Force, April 1995, *IP Over Asynchronous Transfer Mode (IPATM) Working Group Charter*, Internet Engineering Task Force, Reston, VA.

2. ATM Forum Technical Committee, 1996, *The ATM Forum Glossary*, The ATM Forum, Mountain View, CA.

3. Black, U., 1997, *Emerging Communications Technologies - Second Edition*, Prentice Hall PTR, Upper Saddle River, NJ.

4. Goncalves, M. and Niles, K., 1999, *IP Multicasting: Concepts and Applications*, McGraw-Hill, New York, NY.

5. Banikazemi, M., 1997, *IP Multicasting: Concepts, Algorithms, and Protocols*, http://www.cis.ohio-state.edu/~jain/cis788-97/ip_multicast/index.htm.

6. Black, U., 1998, *ATM Volume II: Internetworking With ATM*, Prentice Hall PTR, Upper Saddle River, NJ.

7. Downes, K., 1998, *Internetworking Technologies Handbook*, Macmillian Technical Publishing, Indianapolis, IN.

8. IETF, April 1998, *Real-Time Streaming Protocol, IETF RFC 2326*, Internet Engineering Task Force, Reston, VA.

9. IETF, January 1996, *RTP: A Transport Protocol for Real-Time Applications, IETF RFC 1889*, Internet Engineering Task Force, Reston, VA.

10. IETF Network Working Group, August 1989, *Host Extension for IP Multicasting: IETF RFC 1112*, Internet Engineering Task Force, Reston, VA.

11. IETF Network Working Group, July 1993, *Multiprotocol Encapsulation over ATM Adaptation Layer 5: IETF RFC 1483*, Internet Engineering Task Force, Reston, VA.

12. Stardust Technologies, Inc., February 1997, *Higher Level Protocols used with IP Multicast: An IP Multicast Initiative White Paper*, Stardust Technologies, Inc., Campbell, CA.

13. Sackett, G. and Metz, C., 1996, *ATM and Multiprotocol Networking*, McGraw-Hill, New York, NY.

14. ATM Forum Technical Committee, 1996, *Integrated Local Management Interface (ILMI) Specification - Version 4.0*, The ATM Forum, Mountain View, CA.

15. IETF, November 1996, *Support for Multicast over UNI 3.0/3.1 based ATM Networks, IETF RFC 2022*, Internet Engineering Task Force, Reston, VA.

16. Chappell, L. and Spicer, R., 1994, *Novell's Guide to Multiprotocol Internetworking*, Novell Press, Alameda, CA.

17. Wittman, R., et al, June 1998, "Amnet: Active Multicasting Network," *ICC'98: 1998 IEEE International Conference on Communications*, Institute of Electrical and Electronics Engineers, Atlanta, GA.

18. Xie, Y., et al, June 1997, "Multicasting Over ATM Using Connection Servers," *ICC'97: 1997 IEEE International Conference on Communications, Institute of Electrical and Electronics Engineers*, Montreal, Canada.

19. Johnson, V. and Johnson, M., 1997, *IP Multicast Backgrounder: An IP Initiative White Paper*, Stardust Technologies, Inc., Campbell, CA.

20. Stardust Technologies, Inc., February 1997, *IP Multicast Initiatives*, Stardust Technologies, Inc., Campbell, CA.

21. ATM Forum Technical Committee, July 1996, *ATM User-Network Interface (UNI) Signalling Specification Version 4.0*, The ATM Forum, Mountain View, CA.

22. Freeman, R., 1996, *Telecommunications System Engineering - Third Edition*, John Wiley & Sons, Inc., New York, NY.

23. Fortino, A. and Golick, J., 1996, *Multivendor Networking*, McGraw-Hill, New York, NY.

24. Monday, M., April 1999, "IP QoS: At the Edge and in the Core," *Telecommunications*, Horizon House, Norwood, MA.

ACRONYMS

AAL	ATM Adaptation Layer
ACK	Acknowledgment
AESA	ATM End System Address
AFI	Authority and Format Identifier
ANSI	American National Standards Institute
ATM	Asynchronous Transfer Mode
ATMARP	ATM Address Resolution Server
B-ICI	B-ISDN Inter-Carrier Interface
B-ISDN	Broadband Integrated Services Digital Network
CBR	Constant Bit Rate
CBT	Core-Based Tree
CNAME	Canonical Name
DARPA	Defense Advanced Research Projects Agency
DCC	Data Country Code
DoD	Department Of Defense
DSP	Domain Specific Part
DSU	Digital Service Unit
DVMRP	Distance Vector Multicast Routing Protocol
ESI	End System Identifier
GOSIP	Government Open Systems Interconnection Profile
HEC	Header Error Control
ICD	International Code Designator
IDI	Initial Domain Identifier
IE	Information Element
IEEE	Institute of Electrical and Electronics Engineers
IETF	Internet Engineering Task Force
IGMP	Internet Group Management Protocol
ILMI	Integrated Local Management Interface
IND	Indication
InterNIC	Internet Information Center
IP	Internet Protocol
IPATM	IP Over ATM
IPv4	IP version 4

IPv6	IP version 6
ISDN	Integrated Services Digital Network
ISO	International Standards Organization
ITU	International Telecommunication Union
ITU-T	ITU - Telecommunications Standardization Sector
IWU	Interworking Unit
LAN	Local Area Network
LLC	Logical Link Control
MARS	Multicast Address Resolution Server
MBONE	Multicast Backbone
MCS	Multicast Server
MOSPF	Multicast Open Shortest Path First
MPOA	Multiprotocol Over ATM
MROUTER	Multicast Router
N6	Technology and Standards Division
NANP	North American Numbering Plan
NCS	National Communications System
NIST	National Institute of Standards and Technology
NNI	Network-Network Interface
NSAP	Network Service Access Point
NS/EP	National Security and Emergency Preparedness
OSI	Open Systems Interconnection
OMNCS	Office of the Manager, National Communications System
OSPF	Open Shortest Path First
PDU	Protocol Data Unit
PIM	Protocol-Independent Multicast
PIM-DM	PIM-Dense Mode
PIM-SM	PIM-Sparse Mode
PTO	Public Telecommunications Operator
PVC	Permanent Virtual Circuit
QoS	Quality of Service
RFC	Request For Comment
RPM	Reverse Path Multiplexing
RSVP	ReSerVation Protocol
RTCP	Real-Time Control Protocol
RTP	Real-Time Transport Protocol

RTSP	Real-Time Streaming Protocol
SAP	Service Access Point
SAR	Segmentation and Reassembly
SDU	Service Data Unit
SVC	Switched Virtual Connection/Circuit
TCP	Transmission Control Protocol
TCP/IP	Transmission Control Protocol/Internet Protocol
UDP	User Datagram Protocol
UNI	User-Network Interface
VBR	Variable Bit Rate
VC	Virtual Circuit/Channel
VCC	Virtual Channel Connection
VCI	Virtual Circuit/Channel Identifier
VCL	Virtual Channel Link
VP	Virtual Path
VPI	Virtual Path Identifier
WAN	Wide Area Network