ATM Forum: Technical Committee

ATM Forum/ 95-0824r6

****

Source: Multiprotocol Sub-Working Group

Editor: Caralyn Brown
Bay Networks, Inc.
2 Federal Street
Billerica, MA 01821

Phone: (508) 436-3835
Fax: (508) 670-8760
Email: cbrown@baynetworks.com

****

Title: Baseline Text for MPOA

****

Date: February 26, 1996

****

Distribution: Multiprotocol Sub-Working Group

****

Abstract: This contains the baseline text for the MPOA sub-working group as agreed to at the February technical committee meeting held in Beverly Hills, CA.

****

****

0. Work Topic List

1. Spanning tree - augmentation and clarification of text.
2. Details of component behavior.
    - AHFG
    - general service description
    - ATM host behavior
    - RSFG/RFFG
    - ICFG/DFFG
    - cache management
    - EDFG (Mike See)
        - Wording of how to deal with EDFG looping back
        - abstraction for legacy ports
        - ATM to legacy forwarding behavior (elaboration needed)
    - Mandatory behaviors (Eric Gray and Joel Halpern)
3. Protocol description (NHRP, MARS, LANE references)
    - enhancements for registration
    - enhancements for cache management
    - enhancements for piggy-backed MAC information
    - NHRP (Eric Gray)
4. Day in the life of a packet (Norm Finn and others)
5. Configuration.
    - Failover - standby how do you find them? (James Watt)
    - Dynamics and resiliency of configuration.(James Watt)
    - what is to be configured
    - configuration validity and consistency checks.
    - definition of configuration database and management of same
6. Replication of ICFGs (includes ipmc)
    - load sharing (per protocol and per service area)
    - redundancy
    - fail-over
    - selection of ICFG (see failover and sharing etc.)
    - capabilities (load sharing, some can do things and some cannot)
    - coordination
7. Signaling. (Code points etc. UNI 3.x and 4.x considerations)
8. Multicast issues
9. QoS support/issues.

10. Filtering support/issues.
11. IASG labeling - Partitioning or what tells you how to remerge?
12. External router discovery. (router on an ethernet...reaching through your broadcast? low priority)
13. Discovery/registration of internal components.
14. Support of additional legacy ports - non-LAN ports. (interfaces at the edge of MPOA)
    - Frame Relay
    - Serial lines
    - etc.
15. Elaborate on behavior of non-routable protocol text.

1. Introduction

1.1 State of this Document

Specific base choices were made to provide clear architecture. The detailed design choices in this document are open for modification.

Where applicable and equivalent we will re-use mechanisms from existing standards and not arbitrarily make alternate design choices. In particular, aspects of the LANE mechanisms, either current or prospective, may be applicable to the intra-IASG (Internetwork Address Sub-Group) protocols while aspects of the NHRP mechanisms may be applicable to the inter-IASG protocols.

It is acknowledged that this document may contain design choices that conflict with the foregoing statements and thus may need to be revised.

2. Terms and Definitions.

2.1 Definitions

Direct Set: A set of host interfaces which can establish direct layer two communications for unicast (not needed in MPOA).

Edge Device: A physical device which is capable of forwarding packets between legacy interfaces and ATM interfaces at any protocol layer, including, but not limited to, the internetwork layer It may or may not also participate in running one or more internetwork layer routing protocols.

Emulated LAN: See ATM Forum LANE specification (ATM Forum document AF-0021).

Forwarding Description: The resolved mapping of an MPOA Target to a set of parameters used to set up an ATM connection on which to forward packets.

Host Apparent Address Sub Group:

A set of internetwork layer addresses which a host will directly resolve to lower layer addresses.

Internetwork Address Sub-Group (IASG):

A range of internetwork layer addresses summarized in an internetwork layer routing protocol.

Internetwork Broadcast Sub-Group (IBSG):

	A set of hosts which can receive a particular internetwork layer broadcast. Note that the scope of this is normally the same as IASG.
Internetwork Layer:	e.g. IP, IPX, DECnet routing, CLNP, AppleTalk DDP, IPv6, Vines
LANE Port	An abstraction over LANE clients and ELANs which allows self-determination of LANE behavior. Data frames delivered from an IASG to LANE or from LANE to a IASG are logically associated with a single LANE port.
MPOA:	An effort taking place in the ATM Forum to standardize protocols for the purpose of running multiple network layer protocols over ATM.
MPOA Client:	A protocol entity which implements the client side of the MPOA protocol, , i.e. it exhibits either Edge Device Functional Group (EDFG) or ATM-attached Host Functional Group (AHFG) behavior.[1]
MPOA Server	A protocol entity which implements the server side of the MPOA protocol, i.e. it contains one or more of any of IASG Coordination Functional Groups (ICFGs), Route Server Functional Groups (RSFGs) or Intra-IASG Broadcast/Unknown Functional Groups (IBUFGs).[2]
MPOA Service Area:	The collection of server functions and their clients . A collection of physical entities consisting of an MPOA server plus the set of clients served by that server.
MPOA Target:	A set of {protocol address, path attributes, e.g. internetwork layer QoS, other information derivable from received packet} describing the intended destination and its path attributes that MPOA devices may use as lookup keys.
One Hop Set:	A set of hosts which are one hop apart in terms of internetwork protocol TTLs (TTL=0 “on the wire”).
Router:	A physical entity that is capable of forwarding packets based on internetwork-layer information, and which also participates in running one or more internetwork-layer routing protocols.
Routing Protocol:	A general term indicating a protocol run between routers and/or route servers in order to exchange information used to allow computation of routes. The result of the routing computation will be one or more forwarding descriptions.
Subnet:	The use of the term “subnet” to mean a LAN technology is a historical use and is not specific enough in the MPOA work. Please refer to Internetwork Address Sub-Group, Direct Set, SNDCF, Host Apparent Address Sub-Group and One Hop Set for more specific definitions.

2.2 Addressing Models

Peer Addressing Model:	A model for mapping the addressing of an internetwork layer onto that used by the ATM fabric. It assumes that a sender can locally algorithmically translate the desired internetwork-layer destination address into the correct ATM address to use to place a call to that destination.
Separated Addressing Model:	A model for mapping the addressing of an internetwork layer onto that used by the ATM fabric. It assumes that a sender must perform a dynamic lookup of the mapping from the desired internetwork-layer destination address to the correct ATM address to use to place a call to the destination. This mapping

[1] MPOA Clients and MPOA Servers may be co-located in a single physical entity.

[2] MPOA Servers and MPOA Clients may be co-located in a single physical entity.

information is provided by the destination or some proxy for it (e.g. network management or a network server) either a priori or else in response to a query. [3]

2.3 Routing Models

Integrated Routing Model: A model used to represent the internetwork-layer routing topology on top of an ATM topology. It integrates the internetwork-layer topology with that of the underlying ATM infrastructure into a single database of information.

Layered Routing Model: A model used to represent the internetwork-layer routing topology on top of an ATM topology. It separates information about the underlying ATM infrastructure from the layers above it, e.g. current utilization/congestion reports.[4]

2.4 Routing and Addressing Model Interactions

The definitions of the addressing and routing models are included to assist in the understanding of various configurations. Any combination of routing and addressing models is acceptable. That is, the selection of routing and addressing models is independent.

2.5 Acronyms

AHFG	ATM-attached Host Functional Group
EDFG	Edge Device Functional Group
FG	Functional Group
DFFG	Default Forwarder Functional Group
IASG	Internetwork Address Sub-Group
IASGid	Internetwork Address Sub-Group Identifier
IBSG	Internetwork Broadcast Sub-Group
ICFG	IASG Coordination Function Group
MPOA	Multiprotocol Over ATM
NHRP	Next Hop Resolution Protocol
QoS	Quality of Service
RFFG	Remote Forwarder Functional Group
RSFG	Route Server Functional Group
SVC	Switched Virtual Channel

3. MPOA Service Overview

3.1 Services Provided by MPOA

The fundamental purpose of the MPOA Service is provide end-to-end internetworking layer connectivity across an ATM fabric, including the case where some internetworking layer hosts are attached directly to the ATM fabric and some are attached to legacy subnetwork technologies.

An internetwork can be sub-divided into one or more non-overlapping Internet Address Sub Groups (IASGs). An IASG is defined to be “a range of internetwork layer addresses summarized into internetwork layer routing.” For the purposes of this document, the scope of an IASG is taken to be identical to the Internetwork Broadcast Sub-Group (IBSG) for each of the members of the IASG (this saves creating yet another term, or heaven forbid, yet another acronym). An IASG is internetwork-protocol specific, that is if a host operates two internetwork layer protocols, it is a member of, at least, two IASGs.

---

[3]:This is sometimes also known as “Overlay Addressing.”

[4] This is sometimes also known as the “Overlay Routing Model.”

Within an MPOA System, an IASG is identified by an IASG Identity (often shortened to IASGid). This identity value is an opaque bit pattern of a length and format to be determined; the definition of values for and semantics of a small number of distinguished values, e.g. “black hole”, “null”, “foreign”, is also a matter for further study.

3.2 Services Required by MPOA

[editor’s note: we need to specify the ATM services (or any other services) required to support MPOA]

4. Elements of the MPOA Solution

4.1 Overview

The MPOA solution presented in this document consists of a number of logical components and information flows between those components . The logical components are of two kinds: MPOA clients and MPOA servers – see section 4.2 for details of the components and their categorization . The information flows between logical components, described in Section 4.3 below, are carried over ATM SVCs using LLC/SNAP encapsulation.

MPOA servers maintain “complete” knowledge of the MAC and internetworking topologies for the IASGs they serve. To accomplish this, they must exchange information among themselves and with MPOA clients.

MPOA clients maintain local caches of mappings (from packet prefix to ATM information). These caches are populated by requesting the information from the appropriate MPOA server on an as-needed basis. The clients use these mappings to forward user data packets as required. These caches are aged by the clients, however the lifetime of a cache entry is specified by the functional group that provided it, i.e. ICFG (IASG Coordination Functional Group) or RSFG (Route Server Function Group). The provider of a cache entry may “retract” it at any, e.g. when routing changes or a station moves.

Replication of server components may be required for reasons of capacity and/or availability. The solution described herein explicitly admits and provides for the replication of server comments with one important caveat: clients are blissfully unaware of these mechanisms.

4.2 Logical Components of the MPOA Solution

For clarity, this document attempts to restrict itself to the discussion of functions performed by logical components and defers the composition of logical components into physical devices to those who develop products.

Functional Group (FG): A collection of "functions" related in such a way that they will be provided by a single logical component of MPOA. Examples include the Route Server Functional Group (RSFG) the IASG [Internetwork Address Sub-Group] Coordination Functional Group (ICFG), and the ATM-attached Host Functional Group (AHFG). One or more of any of these functional groups may co-reside in the same physical entity. MPOA allows arbitrary physical locations of these groups.

All of the Functional Groups present in an MPOA System are shown in Figure 1 and described briefly below.

The *Edge Device Functional Group (EDFG)* is the group of those functions that are performed by a device which provides internetworking level connections between legacy internetworking subnetwork technologies and ATM.

The *ATM-attached Host Functional Group (AHFG)* is the group of functions performed by an ATM-attached host that is participating in the MPOA service.

The *IASG Coordination Functional Group (ICFG)* is the group of functions performed to coordinate the physical distribution of a single IASG across multiple legacy LAN physical ports and/or ATM stations, e.g.

permitting the distribution of a single IPv4 subnet across multiple Ethernet segments whilst maintaining correct operation.

The *Default Forwarder Function Group (DFFG)* is the group of functions performed in association with forwarding traffic within an IASG in the absence of direct client to client connectivity. The DFFG may also forward traffic to other forwarders when the traffic is destined outside the IASG in the absence of direct client to forwarding server connectivity. The DFFG also acts as the Multicast Server (MCS) in an MPOA based MARS implementation. The *Route Server Functional Group (RSFG)* is the group of functions performed to provide internetworking level functions in an MPOA System. This includes running conventional internetworking Routing Protocols and providing inter-IASG destination resolution.

The *Remote Forwarder Functional Group (RFFG)* is the group of functions performed in association with forwarding traffic from one IASG to another, from one MPOA client in one IASG to another IASG, from one IASG to the destined MPOA client in another IASG or from an MPOA client in one IASG to an MPOA client in another IASG.

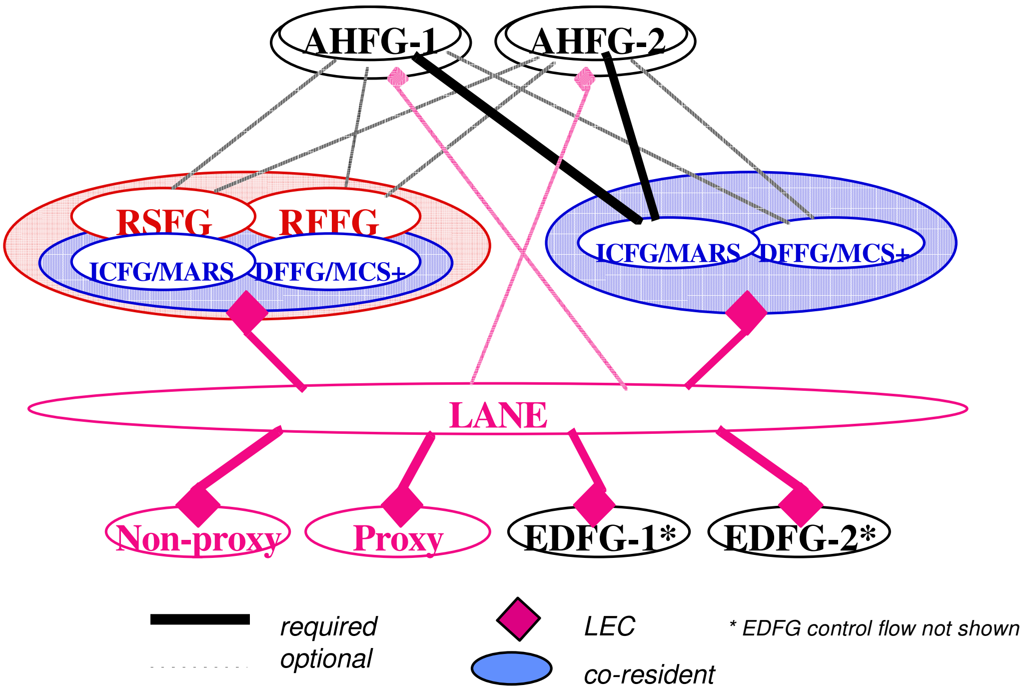


Figure 1 - The Function Groups in an MPOA System

4.3 Information Flows in the MPOA Solution

4.3.1 Summary

The MPOA solution involves many information flows which can be categorized as follows:

a) configuration flows – all functional groups use such a flow to retrieve configuration information,

b) data transfer flows – the end goal of the system (note that communications between an RSFG and a router appear as data transfer to the rest of the system)

c) client-server control flows – used by clients to inform and query the MPOA servers and

d) server to server flows – used by servers to provide the illusion of a single service whilst distributing the service across multiple devices for reasons of capacity and/or availability. Figure 2 shows the client-server and server-server information flows. The data transfer and configuration flows are suppressed to keep the diagram readable.

Section 4.3.4 discusses the encapsulation of data within a flow and Section 4.3.5 discusses issues with the reliable transmission of control and data frames. More detailed discussion of the information flows is contained in the area descriptions in Section 5.2 through 5.7.

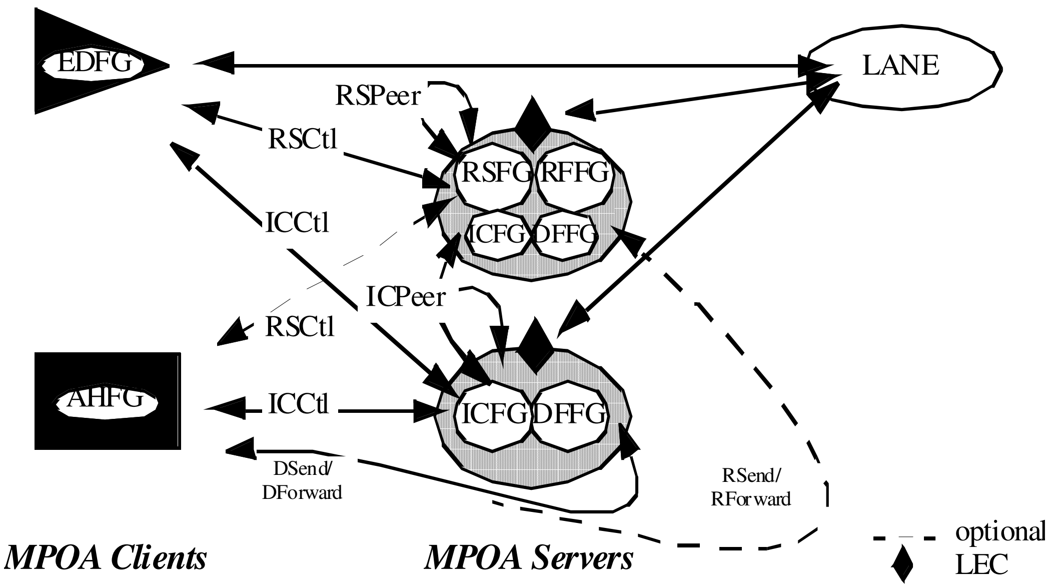


Figure 2 - Information Flows in an MPOA System

Details of SVC establishment and detailed header and control packet formats are both matters requiring further study. The detailed data formats of user data are specified as part of the internetworking protocol used and thus need not be specified as part of the MPOA solution. Issues associated with SVC setup include, but are not limited to: aging of VCCs, provision and use of QoS information, selection of signaling Information Elements (IEs), and response to disconnects.

The *Configuration* information flows are not shown in Figure 2. Each functional group of the MPOA solution, client and server alike, creates such a virtual circuit at startup and uses it to retrieve the necessary configuration information. Configuration is discussed in Section 5.2.

The *Data Transfer* information flows are not shown in Figure 2 above. These information flows are used as described in Section 5.5 to provide ATM-layer connectivity between any combination of EDFGs and AHFGs.

4.3.2 Client to Server Flows

There are six distinct kinds of client-server flows in the MPOA solution: RSFG Control (RSCtl), ICFG Control (ICCtl), Send to DFFG (DSend), Forward from DFFG (DForward), Send to RFFG (RSend) and Forward from RFFG (RForward).

RSCtl information flow is used by MPOA clients to obtain information from the RSFG when trying to resolve the destination internetwork address to an ATM address for inter-IASG data transfer.

ICCtl information flow is used by MPOA clients to obtain information from the ICFG in support of destination resolution in the intra-IASG case and may also be used by the EDFG to register dynamically discovered legacy devices. Legacy device discovery is an optional feature of the EDFG that is discussed in Section 5.3.5 and address resolution is described in Section 5.4.

DSend information flow is used by MPOA clients and servers to transmit data frames in the absence of a direct, client-to-client, short-cut connection and for broadcast and multicast frames. Broadcast and multicast data transfer is described in Section 5.5.2 and unicast data transfer is described in Section 5.5.1.

DForward information flow is used by the DFFG to forward frames to at least those MPOA clients and servers, within the IASG, addressed by the frame.

RSend information flow may be used by MPOA clients (and potentially the DFFG) to forward frames out of an IASG while RForward information flow is used to forward frames into an IASG in the absence of direct IASG to IASG connectivity.

4.3.3 Server to Server Flows

There are two distinct kinds of server to server information flows in the MPOA solution: RSFG to RSFG (*RSPeer*) flows, ICFG to ICFG (*ICPeer*) flows. These flows are labeled in Figure 2.

The *RSPeer* information flow is used by an RSFG to forward Destination Resolution queries (and to receive the response) for destinations in an IASG that is not served by the original RSFG and TBD for replication support.

The *ICPeer* information flow is used by an ICFG to distribute topology information to all of the ICFGs that serve a given IASG. For each ICFG, this flow can be either a single point-to-multipoint or collection of point-to-point VCCs.

4.3.4 Encapsulation

All information transferred between functional groups in the MPOA solution is transferred using the LLC/SNAP encapsulation header defined in RFC 1483; that is, all frames transmitted over ATM by an MPOA device are always transmitted with an RFC 1483 LLC/SNAP header.

The LLC/SNAP code points used by the MPOA solution require specification but will include any existing code points that can be re-used. New code points will be defined using the ATM Forum's OUI. At least two code points are needed: one to indicate those frames that include a MAC header (such frames are considered "for bridging") and (at least) one to indicate those frames that do not include a MAC header (such frames are considered "for routing"). The "for bridging" case is required for those frames where the required MAC header cannot be correctly constructed from the contents of the frame (e.g. IPv4 ARP). Details of the information carried in each type of frame are for further study.

The call setup behavior, the exact details of which require specification, used to establish VCCs among functional groups provides for the two-way negotiation of desired traffic content for a virtual circuit. This allows consenting functional groups to negotiate packet formats per VC. Functional groups may, therefore, choose to override the default LLC/SNAP encapsulation, and use no encapsulation header at all.

The LLC/SNAP encapsulation header permits sharing of virtual circuits between the various transfer mechanisms used by the MPOA solution. The LLC/SNAP header is always transmitted, even over non-shared virtual circuits, to retain the simplicity of encapsulation and driver operations by separating the virtual circuit selection behavior from the encapsulation behavior. In addition to simplicity, this is a more robust design, and we can afford 8 bytes per frame.

The *default*, interoperable behavior supported by all compliant MPOA devices is the sharing of virtual circuits.

4.3.5 Control and Reliability

In the above description, no mention was made of whether individual messages were sent reliably or not. Clearly, since the internetwork protocols being supported are unreliable, the data can be sent unreliably. However, the same can not be easily said of the control messages.

In particular, the EDFG is performing a forwarding function based on information extracted control messages. If this information is sufficiently out of date, black holes or forwarding loops may develop.

The existing time constraints are that the removal of a stations from a bridge forwarding level may not be noticed for 5 minutes, and that bridging topology changes get propagated in 15 seconds, with 15 seconds to recover.

In both these cases, the controlling entity is the EDFG. This apparently simplifies the problem, since it can control the forwarding. To do so, it will at least need a way to inform the ICFG of a forwarding error. As the EDFG examines the destination portion of the MAC header of every “for bridging” packet (see Section 5.5.1), it can detect destinations that are not present on legacy subnetworks directly attached to the EDFG. This also protects intra-IASG transfers from AHFGs to legacy-attached hosts. The notification of a bridging entity move/removal must be done reliably to the ICFG, and reliably from the ICFG to the sources using that information.

For routing, the EDFG is not involved in the detection. The RSFG will determine that routing is changing its path selection. At that time, notification is needed. Reasonable time constraints for routing probably are 15 seconds for black hole repair, and well under 5 seconds for loop repair. These times are shorter than the spanning tree times, since they are driven by a different protocol regime. They are shorter than the actual routing protocol times because they are the reaction times to the changes made by/to routing.

In order to meet these times the update information must be propagated reliably with very short retransmit timers.

The use of short retransmission timers has effects on the interaction with wide area links. Whether we relax the time constraints or accept extra re-transmits of control information on such links is for further study.

4.4 ATM Attached Host Models

editor’s note: further contributions are required to fill in the text of this section. Please align contributions with the motions that were passed during the December meeting.

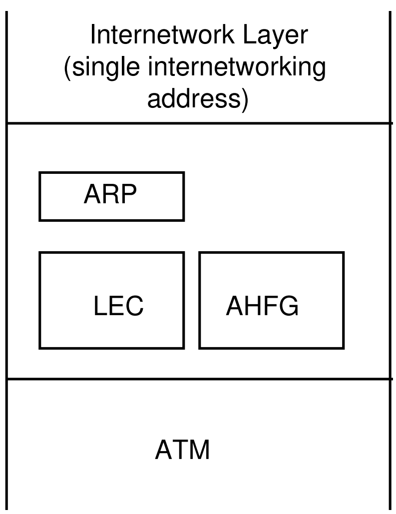


Figure 3 - Dual Mode Host

5. MPOA Service Specification

5.1 Summary

An MPOA System, which is comprised of the functional groups described in Section 4.2, covers the following areas:

a) *Configuration*: this ensures all functional groups have the appropriate set of administrative information;

b) *Registration and Discovery* : this includes the functional groups informing each other of their existence and of the identities of attached devices and EDFGs informing the ICFG of legacy devices,

c) *Destination Resolution*: this is the action of determining the route description given a destination Internetwork Layer address (and possibly other information, e.g. QoS). This is the part of the MPOA System that allows it to perform “cut through” (with respect to IASG boundaries);

d) *Data Transfer:* is getting internetworking layer data from one MPOA client to another;

e) *Intra-IASG Coordination*: is the function that allows IASGs to be spread across multiple physical interfaces;

f) *Routing Protocol Support*: this allows and MPOA System to interact with conventional internetworks;

g) *Spanning Tree Support*: this allows the MPOA System to interact correctly with existing extended LANs;

h) *Replication Support:* this provides for replication of key components for reasons of capacity or availability.

Sections 5.2 through 5.8 provide more detailed descriptions of each of these areas.

As discussed in Section 4.1, the system functions as a set of functional groups that exchange information in order to exhibit the desired behavior(s). To provide an overview of the system’s operation, the behavior of the components is described in a sequence order by “significant events.”

In the descriptions which follow, all externally visible information flows in the MPOA System are assume to be carried in VCCs over the ATM fabric. It may be the case, that for some compositions of the functional groups into physical devices, ATM-level communication between certain functional groups is unnecessary. The impact of this is a matter for future work.

5.2 Startup/Configuration Behavior

5.2.1 Summary

All functional groups will have/know/be configured with a way to contact an appropriate configuration server. The number of servers, synchronization of server databases, and related tasks need to be specified, however it is expected the approach will closely follow, if not simply re-use, the mechanisms specified for the LAN Emulation Configuration Service (LECS). Note that the rest of this document assumes that all of the configurations delivered to the various components are consistent across the system.

In the following, the assertion that the configuration server returns a specific RSFG, ICFG, or IBUFG address does not mean there is only one. It merely means that a given station only needs to talk to one of each kind of server for a given IASG. An alternate design would return the full list of alternate servers to the client, and let the client chose.

If we were to mandate dependence on UNI v4.0, or more accurately signaling support of Leaf-Initiated Joins, we could simply distribute the ATM address of the IBUFG as a configuration item for each IASG in Sections 5.2.2 (the RSFG), 5.2.4 (the EDFG) and 5.2.5 (the AHFG). However, in the absence of signaling support of Leaf-Initiated Joins, the ICFG must tell the IBUFG to add the FG to the appropriate *IBUDistribute* VCC when the functional group registers and there is no need to provide the address as configuration information.

5.2.2 RSFG/ RFFG Configuration

When it starts up, an RSFG/RFFG will get the following configuration information:

a) A list of IASGs it is providing service and for each IASG:

i) the identity assigned to the IASG and

ii) the internetwork protocol for the IASG,

iii) the RSFG's internetwork address on the IASG,

iv) the information needed to join an ELAN, if any, for this IASG;

v) the ATM addresses of the EDFGs if required (see section TBD.)[5]

b) A list of routing protocols to use, and the normal configuration information required for each, including interfaces and peers, as defined by the specification of the routing protocol.

The RSFG/RFFG will establish communication with the ELAN, if any, associated with the IASG.

It will then begin operating the routing protocols according to the configuration information. Actual transmission of routing information will be over the lower layer protocol according to the rules for the protocol family.

5.2.3 ICFG/DFFG Configuration

When it starts up, an ICFG/DFFG will get the following configuration information:

a) A list of IASGs it is providing coordination for, and for each IASG:

i) the identity assigned to the IASG and

ii) the internetwork protocol for the IASG,

iii) Other ICFGs to perform coordination with in support of this IASG.

iv) the information needed to join an ELAN, if any, for this IASG,

v) the address(es) for the RSFG/RFFG pair to use for internetwork forwarding,

vi) the information needed to generate MAC addresses on behalf of AHFGs when necessary;

[5] Editor's note: Need contribution for different cache creation initiation methods, and their supporting information requirements.

vii) the ATM addresses of the EDFGs if required (see section TBD.)[6]

All ICFGs serving an IASG know about all others as this avoids single-points of failure or distinguished behaviors.

The ICFG will establish VCs to each distinct member of the list of peer ICFGs. Note that if the same ATM Address appears in several lists of peer ICFGs, only one VC is needed between the two servers. The ICFG/DFFG will establish communication with the ELAN, if any, associated with the IASG

5.2.4 Edge Device Functional Group (EDFG) Configuration

When it starts up, each EDFG will get the following configuration information:

a) A list of IASGs it is to be a member of, and for each IASG:

i) The IASG identity,

ii) The internetwork protocol supported,

iii) the ATM addresses of the RSFGs, RFFGs, ICFGs and DFFGs if required (see section TBD.)[7]

iv) the MTU for the IASG.[8]

v) the information needed to join an ELAN

b) For each legacy port, a list of IASGs supported on the port;

c) The ELAN(s) to talk to for spanning tree propagation.

The EDFG will establish RSCtl and ICCtl information flows, if appropriate and will establish communication with the ELAN associated with each IASG.

The spanning tree ELAN(s)[9] is used to ensure that an entire MPOA family is participating in a single set of Spanning Tree exchanges.

5.2.5 ATM-Attached Host Functional Group (AHFG) Configuration

When it starts up, each host will get the following configuration information:

a) A list of IASGs it is to be a member of, and for each IASG:

i) The IASG identity,

ii) The internetwork protocol supported,

iii) an ICFG to use for the IASG;

iv) the MTU for the IASG.

---

[6] Editor's note: Need contribution for different cache creation initiation methods, and their supporting information requirements.

[7] Editor's note: Need contribution for different cache creation initiation methods, and their supporting information requirements.

[8] This is for further study. Contributions are requested.

[9] Contributions are requested to clarify the spanning tree propagation.

The host will establish VCs to each distinct ICFG. After the VCCs are setup, the host will proceed with registration as described in Section 5.3 below. Note that use of a protocol-specific auto-configuration capability may require configuration information to be provided to the AHFG beyond that listed above.

Note: whether information for coordinating AHFG and LANE requires further thought. Contributions are requested.

5.3 Registration and Discovery

5.3.1 Summary

Registration in the MPOA solutions is the set of exchanges that are used by the functional groups to inform each other of their existence, capabilities and interests. Discovery is the reporting by EDFGs of legacy-attached devices to the ICFG. These two mechanisms are used by the MPOA solution, along with Routing Protocol Support to ensure that the server components have accurate knowledge of the topology of the directly-attached network both at the MAC and internetwork levels.

5.3.2 RSFG Registration and Discovery

When an RSFG detects an external router, it will register the MAC Address of that router with the relevant ICFG, which will in turn notify the relevant EDFG. This functionality is needed to allow the EDFG to distinguish between locally originated and routed traffic.

Note: This paragraph has problems. We need to rethink this and address the potential problems.

5.3.3 ICFG Registration

As the appropriate ICPeer information flow becomes available, each ICFG will register itself with each distinct peer ICFG listed in its configuration. In each case the registering ICFG will send and receive a list of IASGids served.

5.3.4 EDFG Registration and Discovery

When an EDFG sees legacy traffic from a previously unknown internetworking source address, on a given IASG, it may notify the ICFG. Note that the ICFG will in turn pass on the information to any other ICFGs in the same IASG. This registration function depends on the ability of the EDFG to correctly distinguish "local" traffic from that injected by an external router onto that legacy device.

5.3.5 AHFG Registration

As the appropriate VCCs become available, each AHFG will register itself with each distinct ICFG listed in its configuration. The AHFG will provide a list of (IASGid, internetwork address, and optionally MAC address) tuples to the ICFG. The AHFG must also inform the ICFG whether or not to serve as the AHFG's proxy on an ELAN. If the AHFG does not provide a MAC address and the ICFG has been requested to proxy for the AHFG on an ELAN, the ICFG will provide a MAC address for the AHFG.

When a ICFG receives an AHFG registration, it passes it onto all relevant ICFGs s. Again, if another ICFG registers later, the full list of registered AHFGs is delivered to it. This serves to notify all concerned ICFGs of all ATM attached stations, significantly saving resolution steps later on. Further, the ICFG/DFFG will add the AHFG to the "all clients" DForward flow for the IASG.

5.4 Destination Resolution

When an AHFG wishes to send traffic to an internetwork destination, in both the intra- and inter-IASG cases, it needs to determine the ATM Address to use. This is done by sending a query to the ICFG. Upon resolving the query, the ICFG responds to the AHFG with the ATM Address. If the AHFG uses mask and match behavior, it may send appropriate queries directly to an RSFG.

When an EDFG receives a local address resolution request (e.g. an IPv4 ARP) on a legacy medium, it may determine if it already has a cached answer for that IASG. If it does and the destination is behind another EDFG, it forwards the message to the appropriate destination. If the destination is an AHFG, the EDFG may simply create the proper address resolution response with the MAC address of the AHFG. If it does not have the information, the EDFG forwards the frame to the emulated LAN so that the frame can be forwarded to all other elements in the IASG.

5.5 Data Transfer

(editor's note: This section still need some work.)

Unicast data flow through the MPOA system has two primary modes of operation: the default flow and the short cut flow. Short cut flows are established by the cache management mechanisms. The default flow mechanism provides for data forwarding when ever short cuts do not exist.

5.5.1 Default Unicast Data Transfer

In the default situation the data frame is sent by an MPOA client to an RFFG or DFFG and then forwarded toward the final destination.

When an AHFG originates the packet, it may detect and choose, for internetworking destinations outside the IASG, to send the packet to the RFFG. Otherwise the AHFG sends the packet to the DFFG.

When an EDFG sends the packet in the default situation, it uses LANE for delivery based on the MAC address in the packet.

For a dual mode host originating data, it will use LANE for transmission of intra-IASG traffic and use the AHFG for inter-IASG traffic.

When a packet arrives at an RFFG for an inter-IASG destination it is forwarded by the routing system. When a packet arrives at a DFFG for a registered AHFG, the internetwork protocol packet is forwarded directly to the AHFG.

When a packet arrives at a DFFG from an AHFG for an intra-IASG destination that is not an AHFG, the DFFG will add the appropriate MAC header and use LANE for further forwarding. If the DFFG does not have enough information to build the MAC header, appropriate address resolution is used to get the MAC information.

When a packet arrives at a DFFG for a destination which is not within the IASG, the packet is forwarded to an RFFG.

5.5.2 Short Cut Unicast Data Transfer

When an MPOA client has an internetwork protocol packet to send for which it has an internetworking short cut, the packet is sent over the short cut VC with the appropriate internetworking encapsulation. Broadcast/Multicast Data Transfer

When an AHFG wishes to send a broadcast or multicast frame, it uses the ICFG in a manner similar to the above. Specifically, it sends a query to the ICFG and gets an ATM address.

Inter-IASG broadcast/multicast transfer will presumably be done either by selective splicing onto the point-to-multipoint connection that will be used by the above, or by having another kind of server within the IASG accept the information and forward it to a different point-to-multipoint circuit for inter-IASG transfer.

For EDFGs forwarding broadcast/multicast, the same set of transmissions is required.

5.6 Component Data Transfer Behavior

5.6.1 EDFG Behavior

For forwarding of data frames, the EDFG is assumed to be co-resident with at least one LANE client. Because the relationships between IASGs may be arbitrarily complex (see Figure 4), all data frames delivered from an IASG to LANE or delivered from LANE to an IASG are logically associated with a single LANE port. A LANE port provides an abstraction over LANE clients and ELANs which allows self-determination of LANE behavior independent of this document.

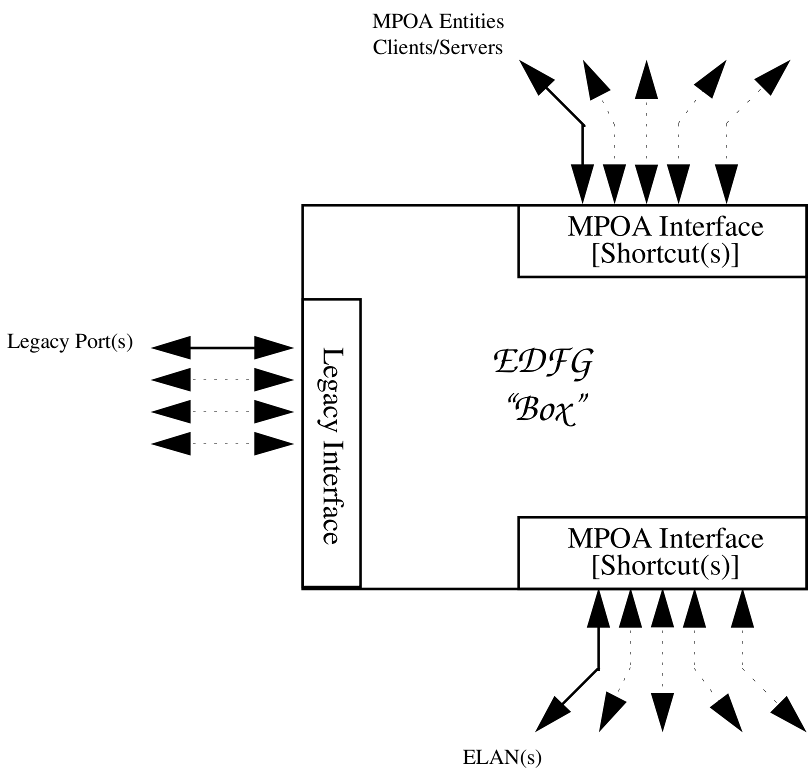


Figure 4 - Example of an Arbitrarily Complex Edge Device Functional Implementation

EDFG behavior (see Figure 5) - relative to both incoming (from legacy to LANE or MPOA) and outgoing (from MPOA to legacy) involves a two-level cache look-up. For incoming traffic, if the destination MAC address corresponds to one for which a cache-table is currently maintained, then the frame's internetwork destination address is read and used as a key to search corresponding table.

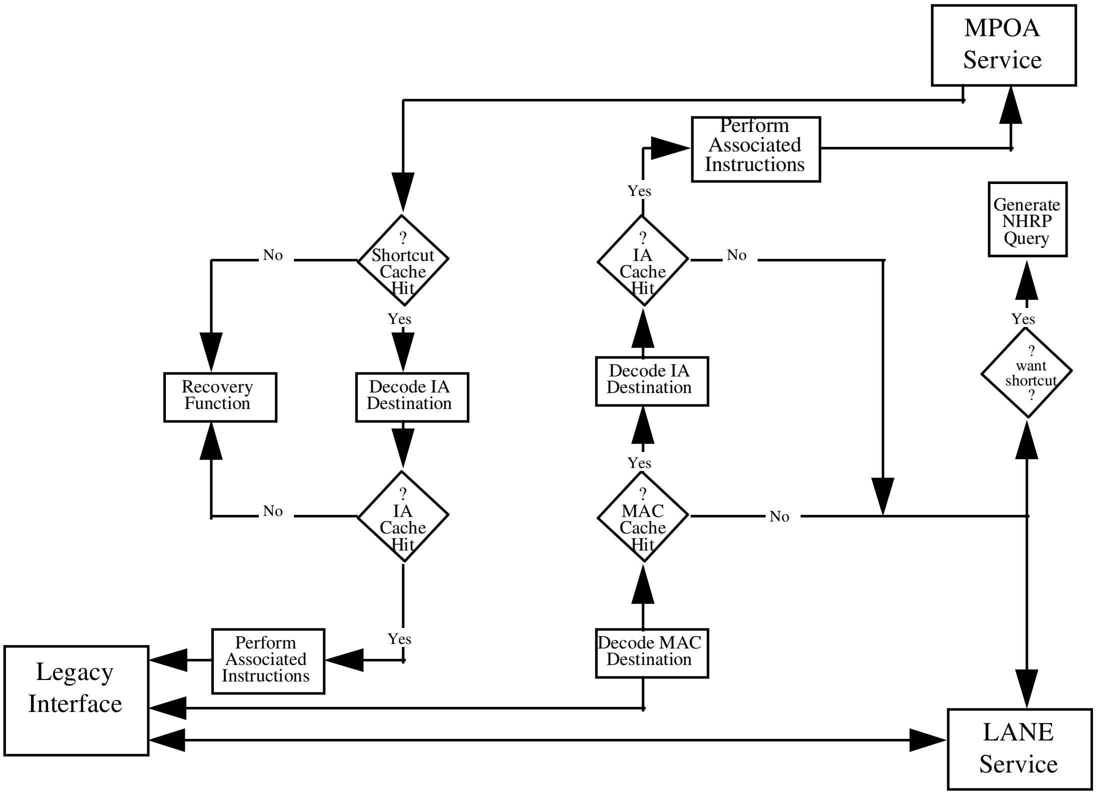


Figure 5 - EDFG Behavior Logical Block Diagram

To perform these lookups, cache tables are assumed to be maintained by the EDFG. For details on assumed characteristics of these cache tables see section 5.6.1.2 below.

A miss on either lookup results in forwarding the traffic via the appropriate LANE port. As a local matter, a miss may also result in origination of resolution protocol request in an effort to establish a cache entry and, potentially, a short-cut data flow. Also as a local matter, it may be necessary to perform an additional check to verify that a data flow is usable (although this may simply be treated as a logical cache miss) and - in the even that it is not usable - processing may be initiated to make it so.

For outgoing traffic, a cache table is sought based on the shortcut data flow on which the data frame arrives. This table is then searched for the destination internetworking address of the frame.

All traffic arriving from a LANE port is forwarded to the appropriate legacy LAN(s) for the associated IASG.

5.6.1.1 Without Cache

In the absence of a cache entry, either because no cache table exists for the destination MAC address, or because a cache entry has not yet been established for this specific MPOA target, the data frame is forwarded on the appropriate LANE port for the associated IASG.

The EDFG may also, as a local matter, determine that a shortcut data flow may be desirable and initiate resolution protocol as described in section 5.4.

5.6.1.2 Cache Management

As mentioned above, the EDFG is assumed to maintain cache tables containing information necessary to determine what legacy LANs, or MPOA shortcuts to use in forwarding data frames. The information expected to be present in each cache entry includes:

- MPOA target

- ATM address of shortcut termination
- Encapsulation information
- IASG ID

Discussion below is intended to provide descriptive explanation of the characteristics of EDFG caching and is not intended to promote or constrain any specific implementation.

Two cache entries are formed from each resolution response (solicited or otherwise): one to be used for data frames arriving from the MPOA system and going to legacy LANs (outbound shortcut cache) and one for data frames arriving from a legacy LAN and going to MPOA (inbound MPOA target cache).

When a resolution response is received, the destination internetworking address, next hop internetworking address, terminating ATM address, source holding time and information relating to encapsulation are used to form an inbound (MPOA target) cache entry. If the EDFG originated the resolution request, it may have already stored information relating the IASG ID and the source internetworking address in an outbound cache entry. If this is not the case, then the EDFG may either resort to locally available information or use legacy address resolution mechanisms to relate the source internetworking address in the response to an IASG ID and corresponding legacy LAN(s). In any case, the source internetworking address in the response is used as the destination internetworking address in an outbound cache entry and stored with an IASG ID and encapsulation information in an outbound cache table. This table should be related to the shortcut data flow associated with this response once it has been established.

Aging of inbound cache entries uses the source holding time from the latest resolution response received relative to the associated destination internetworking address but cache entries may be further limited in duration by management/configuration established maximum holding times. In addition, inbound cache entries may be withdrawn by the originating resolution response authority at any time.

Because outbound cache entries are used exclusively for delivery of data frames to legacy LANs, they are not affected by either source holding time or withdrawn resolution responses and aging of these entries is a local matter. Note that the encapsulation information associated with any particular outbound cache entry and shortcut combination becomes invalid if the shortcut data flow is removed.

editor's note: there is need for further elaboration of the 4 caches discussed in these sections.

5.6.1.3 With Cache - Legacy to ATM

If data frames arriving on a legacy interface result in a cache hit and the shortcut data flow is usable, the encapsulation information in the cache entry is used to de-encapsulate/re-encapsulate the data frame and the frame is then forwarded via the indicated data flow.

editor's note: The case where the EDFG is both the source and the destination of routed traffic needs elaboration. A contribution is needed.

5.6.1.4 With Cache -ATM to Legacy

If data frames arriving from a shortcut data flow result in a cache hit, the encapsulation information in the cache entry is used to de-encapsulate/re-encapsulate the data frame and the frame is then forwarded to the indicated legacy LANs.

editor's note: There is still some question on the need for an abstraction for EDFG to legacy flows. A contribution is needed.

5.6.2 ATM Host Behavior

5.6.2.1 AHFG Behavior

5.6.2.2 Dual Mode Host Behavior

5.6.3 ICFG/DFFG Behavior

5.6.3.1 Intra-IASG Coordination

Each ICFG is given a list of all other ICFGs supporting a given IASG. It establishes a pt-to-multipoint VC to them, or uses a collection of point-to-point VCs. In either case, whenever an ICFG accepts a registration, it sends the information to all other relevant ICFGs. When a registration is removed (time-out, VC termination, ...) all other relevant ICFGs are notified of that fact as well. Additionally, port status changes (see Spanning Tree below) are also passed on.

When an ICFG gets a query from an AHFG which for a destination not within the IASG, it will pass the query to an RSFG registered as serving the source IASG. (It is for further study to determine whether, if an ICFG is also serving the destination IASG, it can safely deliver the answer without consulting the RSFG.)

When an ICFG receives a legacy station registration from an EDFG, it passes that information to all registered relevant RSFGs, so as to avoid the need to query the ICFG. If we wish to eliminate this, then there will be a query message from the RSFG to the ICFG to get the ATM Location, (port information?) and Mac Address for a destination within an IASG.

An ICFG will record which EDFGs have been given what information within the IASG, and whenever such information becomes invalid, the EDFG will be notified to remove the cache entry

5.6.4 RSFG/RFFG Behavior

The RSFG engages in the operation of traditional routing protocols. These may include operation over legacy media reached through EDFG, operation over ATM with identified peers, and operation over ATM Groups using a mesh of point-to-multipoint circuits. The ATM peer operation may be mediated by an IASG/ICFG, or may simply be via configured, numbered or unnumbered, adjacencies. Point-to-multipoint direct operation would be by configuration, possibly with support from a "multicast client registration function" somewhere. Interaction with legacy media would use LANE.

EDFGs may and ICFGs will send queries to an RSFG. If the destination is in another IASG handled by the RSFG, it will generate the response, and record the fact of the response. If the information is later invalidated, a notification will go to the source of the query. (Due to the absence of a direct VC to an AHFG, the notification to that will go through the support ICFG.) A destination may become invalid either because the actual host moved/expired, or due to a routing change.

If the destination is apparently reachable through another RSFG on the ATM, rather than in a directly served IASG, then the query will be passed to that RSFG. When the reply is received, state for detecting routing changes is saved, and the reply passed to the original source of the query. In certain scenarios, it may be the case that distributing the information among the RSFGs is an acceptable replacement for the query forwarding mechanism – this requires further study.

If an RSFG receives a query from another RSFG, it checks if it serves the destination IASG or otherwise directs traffic for the destination away from the ATM Fabric. If instead routing points at another RSFG (or indistinguishably an ATM Router) then it passes the query on (and later passes the response back). If it is responding to the query, it saves enough state to notify the relevant AHFG or EDFG if a relevant routing change occurs, and generates the response back to the asking RSFG.

When a routing change occurs, an RSFG checks if it affects any cache management entries. If so, it notifies the source and/or destination EDFG/AHFG that the cache entry for the given destination is invalid. If the “destination” EDFG was notified, and there is still a VC from the source EDFG/AHFG, then a message will go to that source notifying it to invalidate the VC. This is required to avoid routing loops in the presence of routing information. It is insufficient for certain known topologies, and further analysis in this area is needed.

At this time, all answers to queries will be strictly for the requested host. Future extensions to prefix query/response and management will require care in the interaction with routing aggregation, variable subnetting, and routing policies.

5.7 Spanning Tree Support

Just to re-iterate from a previous contribution, each EDFG must running IEEE 802.1(d) spanning tree on each port. There is a specific IASG just for propagating spanning tree BPDUs among the EDFGs. Each EDFG is responsible for the generation of BPDUs into that IASG, and onto its attached legacy media.

Each EDFG will also notify the ICFG when any of its ports enter the Blocking or Forwarding state. (Note that the "ports" for this purpose include the ATM port, which can become blocked.) The blocking notification is critical since that results in invalidating routing/location information. The ATM Port blocking also interacts with "knowing" where things are, since other EDFGs are not directed to send intra-IASG traffic to that EDFG. The forwarding state notification serves to counterbalance the blocking notification (thus avoiding "assumptions"). In addition, it permits the ICFG to determine if an entire IASG has gone blocked for access from ATM to Legacy. It is for further study whether the ICFG should then re-enable one port for routing, or whether sufficient information should be passed to the RSFGs to permit them to do so. (In the ICFG case, although there are multiple ICFGs, as long as they all come to the same conclusion about which port of which EDFG should be partially re-enabled, only one ICFG will actually do anything.)

The question whether distinct spanning trees are required to support both IEEE 802.1(d) Spanning Tree and Token Ring Spanning Tree simultaneously or whether a single spanning tree will suffice requires further study.

5.8 Replication Support

[editor’s note: This was specified in the summary section, but no text has been suggested]

5.9 Multicast Support

The first version of the MPOA multicast service will focus on providing specific support for multicasting among MPOA clients that are directly attached to an ATM network. Furthermore, the first version will specifically support only intra-IASG multicasting. Inter-IASG multicasting will initially utilize conventional layer 3 inter domain multicasting mechanisms.

An alternative expression of this focus is that it covers the propagation of layer 3 multicast traffic among EDFGs and AHFGs belonging to the same IASG. The MPOA multicast service model assumes that within EDFGs and AHFGs there will be components specifically designed to support the multicasting needs of the functional group as a whole. Further, the MPOA multicast service model assumes that ICFGs will contain components specifically designed to support the multicasting needs of the IASG they serve.

The following derived terms are introduced to identify these components of existing functional groups:

ICFG.mars

Those elements of the ICFG that support the address mapping and group membership traffic requirements of the MPOA multicasting service. Based on the IETF's Multicast Address Resolution Server (MARS) [1].

AHFG.mclient

Those elements of the AHFG that utilize the ICFG.mars to establish and manage ATM resources for intra-IASG multicasting.

EDFG.mclient

Those elements of the EDFG that utilize the ICFG.mars to establish and manage ATM resources for intra-IASG multicasting.

DFFG.mcs An entity within the MPOA multicast service that behaves as a Multicast Server. It has no functions relating to unicast traffic.

The term 'cluster' will be used to indicate the subset of MPOA clients within an IASG that are directly ATM attached. A cluster may contain any combination of AHFGs and EDFGs.[10]

The short-hand term 'mclients' will be used to generically refer to AHFG.mclients and/or EDFG.mclients in the following sections.

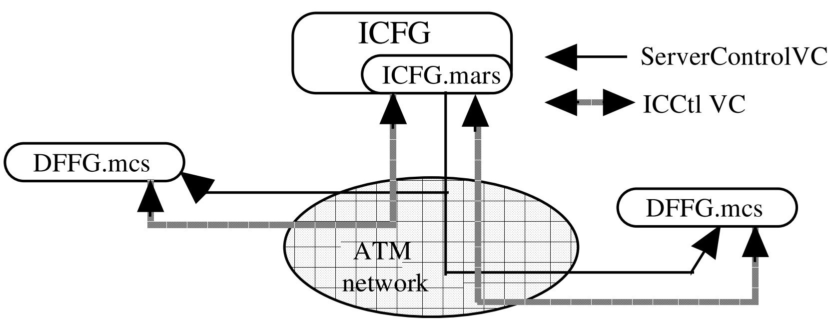
Finally, in this first version of the MPOA multicast service it will be assumed that only one ICFG.mars is actively supporting a given IASG at any time (although multiple ICFG.mars' may be monitoring the IASG activity and be able to take over as backup if the active one fails).

5.9.1 Overview of ICFG.mars and mclient Interaction.

An ICFG.mars serves one or more mclients. Within a given IASG the relationship between the ICFG.mars and its mclients is shown in Figure 6. Each mclient establishes its own bi-directional point-to-point VC to the ICFG when it needs to query the ICFG.mars or register/deregister a group membership for itself. This VC may be transient, and represents an instance of the ICFG Control (ICCtl) VC identified in section 4.3.1.

To support the requirement that mclients are updated quickly with group membership changes that occur in other mclients within the cluster, the ICFG.mars manages a permanent, unidirectional point-to-multipoint VC, ClusterControlVC [*for want of a better name*], out to all cluster members.

ICFG.mars messages on ICCtl and ClusterControlVC are LLC/SNAP encapsulated. They use a unique code-point allocated by the IETF to the MARS protocol. Other protocols may share these control paths between the ICFG and EDFG/AHFGs (e.g. unicast MPOA Target resolution based on NHRP) by using different LLC/SNAP code points.



[10] Editor's note: The term cluster is discussed in the Multicast appendix which follows this section. We need to work on the order of the document and/or define cluster elsewhere.

Figure 6 - Control paths between ICFG.mars and mclients.

All ICFG.mars messages transmitted on ClusterControlVC carry a 32 bit sequence number (Cluster Sequence Number, or CSN). This is incremented for every transmission on ClusterControlVC, and is tracked by mclients to allow them to detect if they've missed a previous general transmission from the ICFG.mars.

5.9.1.1 Message types for ICFG.mars interactions.

The following message types are taken from the MARS protocol. For convenience the IETF mnemonics will be adopted.

MARS_JOIN,
MARS_LEAVE.

Allow endpoints to join and leave specific layer 3 multicast groups. Also used to pass on to the entire cluster asynchronous updates of a group membership change.

MARS_REQUEST.

Allow endpoints to request the current membership list of a layer 3 multicast groups.

MARS_MULTI.

Allows multiple ATM addresses to be returned by the ICFG.mars in response to a single MARS_REQUEST.

MARS_NAK.

Explicit negative response returned by the ICFG.mars if no information is available to satisfy a previous MARS_REQUEST.

MARS_REDIRECT_MAP.

Allow ICFG.mars to specify a set of ATM addresses of backup ICFG.mars entities for its mclients.

MARS_GROUPLIST_REQUEST,
MARS_GROUPLIST_REPLY.

Allow ICFG.mars to indicate which groups have actual layer 3 members. (May be used to support higher layer functions such as IGMP in IPv4 multicast routing.)

Each message carries a 56 bit protocol identifier and an arbitrary byte string (up to 255 bytes long) representing layer 3 protocol multicast group address(es). The first 16 bits of the protocol identifier is an unsigned integer value taken from the following number space:

0x0000 to 0x00FF Protocols defined by the equivalent NLPIDs.
0x0100 to 0x03FF Reserved for future use by the IETF.
0x0400 to 0x04FF Allocated for use by the ATM Forum.
0x0500 to 0x05FF Experimental/Local use.
0x0600 to 0xFFFF Protocols defined by the equivalent Ethertypes.

If the NLPID value of 0x80 is specified, the actual protocol being carried is identified by a SNAP value encoded in the remaining 40 bits of the protocol identifier field. Otherwise, the protocol is entirely specified by the initial 16 bit value.(As examples, a protocol identifier of 0x800 would indicate IPv4, and its multicast addresses would be 4 octets long. A protocol identifier of 0x86DD would indicate IPv6, and its multicast addresses would be 16 octets long.)

The message format is described in more detail in section 5.9.5.

5.9.1.2 Registration.

In order to be added to ClusterControlVC, and so utilize an ICFG.mars, mclients must register with the ICFG.mars for their IASG. This is achieved by exchanging with the ICFG.mars a special form of the MARS_JOIN message on ICCtl VC. Registration also results in the ICFG.mars allocating each mclient a 16 bit ClusterMemberID (CMI, valid only for the time that the mclient remains an active leaf on ClusterControlVC).

A special form of the MARS_LEAVE message is used to deregister a cluster member.

5.9.1.3 Joining and Leaving Groups.

An mclient is a 'group member' (in the sense that it receives packets directed at a given layer 3 multicast group) when its ATM address appears in the ICFG.mars' mapping table entry for a given group's multicast address. It adds its ATM address to the membership map of a given group by sending the MARS_JOIN message on ICCtl VC.

Conversely, to leave a group an mclient needs to remove its ATM address from the ICFG.mars' mapping table for the given group. This is achieved by sending the MARS_LEAVE message on ICCtl VC.

When the ICFG.mars determines that all cluster members need to be informed of a change in group membership it propagates an appropriate MARS_JOIN or MARS_LEAVE on ClusterControlVC.

5.9.1.4 Providing Redundant ICFG.mars Entities.

Every mclient keeps a table of one or more ATM addresses representing the ICFG.mars' it should attempt to register with. This table is sorted in order of preference, and is stepped through sequentially during registration until registration succeeds. This table is also stepped through sequentially if re-registration occurs due to failure of the current ICFG.mars. As a minimum, each mclient must have at least one ATM address in this table.

The table is dynamically updated by the contents of MARS_REDIRECT_MAP messages that are regularly transmitted on ClusterControlVC. These messages contain a list of ATM addresses representing the set of ICFG.mars' that the current ICFG.mars considers to be its backups. When a client receives a MARS_REDIRECT_MAP it copies its contents into its local ICFG.mars table.

Hot swapping of mclients from one ICFG.mars to another is also triggered by MARS_REDIRECT_MAP. These procedures are discussed further in later sections.

5.9.2 Overview of the Data Path Management.

When an AHFG or EDFG requires an outbound path to the members of a layer 3 multicast group, and no such path pre-exists, its mclient constructs a MARS_REQUEST (representing the multicast MPOA Target) and sends it on ICCtl to the ICFG.mars (re-establishing ICCtl too, if necessary).

This query elicits a multicast Forwarding Description from the ICFG.mars in the form of one or more MARS_MULTI messages. The multicast Forwarding Description contains a set of one or more ATM addresses, which the mclient uses to specify the leaf nodes of a point-to-multipoint VC. The mclient then considers this outgoing VC to be the appropriate forward path for multicast traffic to the layer 3 group identified in the original MARS_REQUEST.

If no members exist for the group identified in the MARS_REQUEST, the ICFG.mars returns a MARS_NAK instead.

Once a forward path has been established, the mclient monitors incoming MARS_JOINs and MARS_LEAVEs. These will arrive on ClusterControlVC when the ICFG.mars is informing the entire cluster of a new member's arrival or an old member's demise. If a MARS_JOIN arrives indicating a new member on a group for which the mclient has an active forward path too, a new leaf node is immediately added to the forward path. If a MARS_LEAVE arrives indicating the loss of a member on a group for which the mclient has an active forward path too, the associated leaf node is dropped.

The mclient makes no assumption that the ATM addresses provided by the ICFG.mars in MARS_MULTI or MARS_JOIN messages actually represent the group members. The only assumption is that they represent the forward path determined by the ICFG.mars for that group. This allows the ICFG.mars to introduce multicast servers (DFFG.mcses) into the forward path in a manner transparent to the mclients. If a layer 3 group is being served by an DFFG.mcs then the ICFG.mars returns the ATM address of the DFFG.mcs in the MARS_MULTI. When a new group member joins or leaves an DFFG.mcs supported group, the ICFG.mars informs the DFFG.mcs rather than the mclients.

Data packets transmitted on the forward path are LLC/SNAP encapsulated. *For the purpose of this first iteration* they shall use the IETF's extended encapsulation for multicast packets, which carries within it the CMI allocated during mclient registration. This allows each mclient to detect reflected AAL_SDUs in a layer 3 protocol independent manner.

5.9.3 Overview of ICFG.mars and DFFG.mcs Interaction.

When an DFFG.mcs is configured to support a given layer 3 multicast group (or set of groups) it interacts with the ICFG.mars in an analogous manner to mclients. Within a given IASG the relationship between the ICFG.mars and its DFFG.mcses is shown in Figure 7. Each DFFG.mcs establishes its own bi-directional point-to-point VC to the ICFG, and uses this to query the ICFG.mars for a group's membership, or register/deregister its willingness to support a given group. This VC may be transient, and represents an instance of the ICFG Control (ICCtl) VC identified in section 4.3.1.

When an DFFG.mcs is supporting a group's traffic it requires rapid updates of group membership changes. To support this the ICFG.mars manages a permanent, unidirectional point-to-multipoint VC, ServerControlVC [*for want of a better name*], out to all DFFG.mcses. ServerControlVC is independent of ClusterControlVC.

Messages on ICCtl and ServerControlVC are LLC/SNAP encapsulated. They use a unique code-point allocated by the IETF to the MARS protocol.

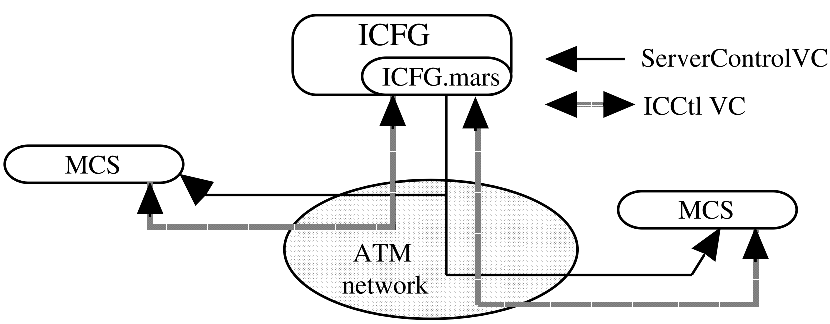


Figure 7 - Control paths between ICFG.mars and MCSs.

The MARS_REQUEST, MARS_MULTI, and MARS_REDIRECT_MAP messages are used by DFFG.mcses in a similar way to mclients.

The following messages are specific to the interactions between an ICFG.mars and DFFG.mcses.

MARS_MSERV,
MARS_UNSERV.

Allow multicast servers to register and deregister themselves with the ICFG.mars.

MARS_SJOIN,

MARS_SLEAVE.

Allow ICFG.mars to pass on group membership changes to multicast servers.

The detailed behavior of the ICFG.mars when handling groups supported by DFFG.mcses is dealt with in a later section.

5.9.4 MPOA Multicast Related LLC/SNAP Code Points.

The LLC/SNAP encapsulation used for ICFG.mars control messages is the same as that used by the IETF's MARS and ATMARP protocols.

```
      [0xAA-AA-03][0x00-00-5E][0x00-03][ICFG.mars control message]
(LLC)        (OUI)      (PID)
```

The extended encapsulation for intra-cluster multicasting of data packets is indicated by another LLC/SNAP codepoint borrowed from the IETF's MARS specification.

```
[0xAA-AA-03][0x00-00-5E][0x00-01][Extended Layer 3 packet]
    (LLC)        (OUI)      (PID)
```

The extended layer 3 packet is encoded in the following manner:

```
[pkt$cmi][pkt$pro][Original Layer 3 packet]
```

The first 2 octets (pkt$cmi) carry the Cluster Member ID (CMI) assigned when an endpoint registers with the ICFG.mars. The second 2 octets (pkt$pro) indicate the protocol type of the packet carried in the remainder of the payload. This is taken from the set of protocol identifiers identified in section 5.9.1.1.

When carrying packets belonging to protocols that require the SNAP form of identification the pkt$pro field is extended to carry the 40 bit SNAP field and padding for 32 bit alignment. Such layer 3 packets are encoded in the following form:

[pkt$cmi][0x00-80][mar$pro.snap][padding][Original Layer 3 packet]
2octet 2octet 5 octets 3 octets N octet

For example, an IPv4 packet would be transmitted as:

```
[0xAA-AA-03][0x00-00-5E][0x00-01][pkt$cmi][0x800][IPv4 packet]
```

As direct multicast packet transmission is currently only intra-cluster, no IASG specific identification needs to be carried on a per packet basis. However, future developments may lead to a new extended packet encapsulation format. These will be identified by a different LLC/SNAP codepoint, and may share forward path VCs with data packets encapsulated using the current scheme.

5.9.5 General Control Message Format.

ICFG.mars messages may further be subdivided.

```
[MARS header][Layer 3 and/or ATM addresses][Suppl. Parameters]
```

[MARS header] contains common fields indicating the operation being performed (e.g. MARS_REQUEST) and the layer 3 protocol being referred to.

The format of the following [Layer 3 and/or ATM addresses] area in the MARS message depends on the operation indicated in the [MARS header]. These provide the fundamental information that the registrations, queries, and updates use and operate on.

Finally, [Suppl. Parameters] represents an optional set of supplementary parameters to be associated with the [Layer 3 and/or ATM addresses] information. Examples of supplementary information would be authentication information, QoS 'hints' to associate with a particular mapping, or encapsulation specifications for data VCs created as a consequence of a particular mapping.

Supplementary parameters are intended to act as modifiers of the default behavior(s) associated with the operation indicated in the header. They need not be supplied if the default rules governing the interpretation of the [Layer 3 and/or ATM addresses] information is sufficient at any given time.

The combination of address fields and supplementary parameters allows a range of multicast MPOA Targets and Forwarding Descriptions to be represented.

6. Detailed Component Behaviors

(Editor's note: we need details here and a better title. Likely this will be several sections)

6.1 Cache Management

(Editor's note: This was a topic of conversation and it was agreed that this issue needs to be addressed)

7. Detailed Protocol

(Editor's note: we need details here and a better title)

8. Signaling Details

(Editor's note: perhaps this is not a separate section, but we need to think about this)

9. References

[1] Grenville Armitage, "Support for Multicast over UNI 3.1 based ATM Networks.", INTERNET DRAFT <draft-ietf-ipatm-ipmc-06.txt>, IP over ATM Working Group, August 1995.

Annex A - NHRP Assumptions

1. Hold down timers must be present in and apply to resolution responses and other protocol messages which may be used to construct cache entries.
2. At least one ICFG s required to be co-resident with each RSFG for each IASG served by that RSFG. (Means that don't have to have an ICFG to RSFG protocol).
3. Any resolution request which needs to be forwarded and which originated at an EDFG must be stashed and re-originated by the RSFG serving the request.
4. All resolution responses are "authoritative"
    - however, not all resolutions responses are "stable
5. Caching of resolution responses at intermediate systems is not allowed.
6. MPOA resolution protocol must provide shortcuts within a single routing domain.
7. MPOA resolution protocol may provide shortcuts within multiple routing domains under certain circumstances (TBD).

(Note: This section needs more text and explanation or needs to be moved into the body of the document)

Annex B - Source-Routing Support in an MPOA Service

B1. Overview of Source Route Bridging

The general format of a token-ring frame is shown in Figure 1. A major difference between LANs that support source-routing such as token-ring and other LAN types is the optional inclusion of the routing information (RI) field (up to 32 bytes) after the destination and source addresses

SD	AC	FC	DA	SA	RI	Payload includnig LLC	FCS	ED	FS

Figure 1: Token-ring frame format

The RI field defines the path a frame will take through the interconnect network of bridges and ring segments. The RI field is built by the end system originating the frame and it is processed by the bridges attached the network to determine if and how the frame should be forwarded. The first two bytes of the RI field is a control field that indicates the following:

- type of explorer frame - all routes or spanning tree only
- length of the RI field
- direction in which the RI field is to be processed
- largest frame that can be sent along the path defined by the RI field

The remainder of the RI field, up to 30 bytes, is composed of 2 byte route designator (RD) fields. The RD field consists of a ring number portion and a bridge number portion. The resulting sequence of ring numbers and bridge numbers define a unique path through the bridged network. The bridges in a SRB network examine the RD field to determine if the frame should be forwarded by the bridge. In the case of explorer frames, a bridge adds its bridge number and the next ring segment to the RI field and increases the length of the RI field as the frame traverses the network.

A bridge or other device supporting source route bridging must support three modes of forwarding frames: specifically routed frames (SRF), spanning tree explorer (STE) frames and all routes explorer (ARE) frames. Specifically routed frames contain a complete RI field that the bridge must examine to determine if it must forward the frame. Explorer frames are used by end systems to discover and build the RI field. Spanning tree explorer frames, as the name implies, are only forwarded by those bridges that are part of the spanning tree created by the bridges. All routes explorer frames are forwarded by all bridges in the network unless the frame has already traversed the next ring segment in the network. All routes explorer frames allow an end system to discover if multiple paths exist to the destination. The end system can then choose the path most suited to its needs: frame size, delay, number of hops, etc.

An important characteristic of SRB networks is that the presence of the RI field allows multiple end systems with the same MAC address to exist in the same bridged network. The only restrictions are that duplicate MAC addresses reside on separate ring segments and that the end systems with the duplicate MAC addresses provide the same service.

B2. SRB Considerations

In order to provide support for source route bridging, some additional function needs to be added to the ICFG/DFFG components. Specifically, a mechanism is needed to obtain source routing information for AHFGs. The RI field is used to define a specific path through the bridged network. Additionally, for non-routable protocols, the RI field is part of the "network" header that should be included in all frames. Also, a mechanism is needed to respond to explorer frames, and receive specifically routed frames on behalf of AHFGs that are not capable of doing so themselves.

B3. Source Route Bridging Support in MPOA

An ATM-attached host with both MPOA and LANE stacks (i.e. with an AHFG and an LEC) may simply use its LEC to leverage LANE functions defined for source routing (note: the same applies to an ATM-attached host with only a LANE stack; however, this is outside the scope of MPOA). However, in order for the host's AHFG to participate in functions that involve source routing, MPOA procedures are necessary. The same procedures are applicable when an ATM-attached host with only an AHFG requires source routing support. Additionally, an EDFG must support cut through connections to/from AHFGs and other EDFGs on which it is receiving and sending frames in the internetwork format. In the following discussion, we identify the situations in which source routing support is required, and outline the procedures for providing this support.

B3.1. AHFG and ICFG/DFFG Behavior

When an AHFG needs to send data to a device on a source route bridged network, a mechanism is needed to obtain the RI field to include in the frames' headers. This mechanism must be part of the MPOA address resolution process. The following actions occur when an AHFG has a frame to send:

- When an AHFG wishes to send traffic to an internetwork destination, the AHFG may directly send to the DFFG the packet that it wishes to transfer to the internetwork destination. It is the ICFG/Duff's responsibility to perform address resolution procedures on behalf of the AHFG in order to obtain the MAC header to prepend to the packet. Alternatively, the AHFG may send an MPOA address resolution request to the ICFG in order to determine the ATM address to use. (Note: if it is determined that the internetwork destination is outside the IASG, the AHFG may send its query directly to the RSFG. However, in that case, source routing support is not required.)
- If the ICFG is able to locate the destination address in its (MPOA) tables, the ATM address of the destination is returned to the AHFG. If the ICFG is unable to find the destination internetwork address in its (MPOA) tables, and the destination is within the IASG, the MPOA address resolution request must be converted to a internetwork protocol address resolution packet (e.g. IP ARP), and broadcast to all members in the IASG. The ICFG/DFFG does this by broadcasting the packet on all the ELANs in the IASG using LANE procedures.
- If it is on a 802.5 LAN, the ICFG/DFFG may receive multiple responses to this query.
- The ICFG/DFFG processes all responses from the target system, and selects the RI field from one of these responses to use in the MAC header of the frames that it will forward on behalf of the AHFG.
- If the protocol supported by the IASG is a non-routable protocol that has been promoted to internetwork status, then the RI field is returned to the AHFG for inclusion in frames as part of the "internetwork address of the protocol". If the AHFG has sent an MPOA address resolution request to the ICFG, then the ICFG returns the ATM address of the ICFG/DFFG to the AHFG. The AHFG establishes a VC to the ICFG/DFFG and begins sending data. The ICFG/DFFG will forward frames from the AHFG to the target after including the MAC header in them. The ICFG/DFFG will use LANE to forward these frames.

The ICFG/DFFG is also responsible for responding to explorer frames for the AHFGs for which it is serving as a proxy on an ELAN.

B3.2 EDFG Behavior

The EDFG behavior in support of SRB networks must be in accordance with the IEEE 802.1 specification [C2] and the LAN Emulation specification [C1]. This includes assignment of bridge numbers and ring numbers, and processing of explorer frames and specifically routed frames. This behavior does not need to be addressed by MPOA.

Additionally, an EDFG must support cut through connections to/from AHFGs and other EDFGs on which it is receiving and sending frames in the internetwork format. When the legacy network is supporting source routing, the EDFG is responsible for maintaining an RI field that is included as part of the MAC header sent on the legacy network.

On frames received from the legacy networks, the EDFG strips the MAC header including the RI field, unless the RI field is part of the internetwork address, before sending the frame on the MPOA direct connection.

Editor's note: this should eventually be folded into the main text.

B4. References

[C1] Ellington, B. (editor), "LAN Emulation Over ATM: Version 1.0 Specification," ATMF/94-0035R9, January 1995.

[C2] ISO/IEC 10038; ANSI/IEEE Std. 802.1d, "Information Processing Systems - Local Area Networks - (MAC Bridges).

Appendix A - Issues Surrounding MPOA Support for Layer 3 Multicasting

A1. Multicast Issues Introduction

In addition to unicast, MPOA will have to address the need for multicast and broadcast services by layer 3 data protocols. This contribution is intended as an introduction to the sorts of issue we will face in this area. Specification of actual solutions will occur in further contributions later this year. The ideas in this contribution owe a lot to the discussions held on the IETF's IP-ATM working group mailing list over the last 10 months. An overview is provided of this related work in the IETF', along with pointers on how to access their latest documents on multicast over ATM.

The intention of the MPOA working group is to provide multicast information by referencing the IETF's multicast RFC when it is completed. If the working group determines that this RFC is insufficient to meet the needs of MPOA, the group will raise the issues at the IETF and pursue a joint solution. In the case where differences remain, this document will reference the RFC and point out differences (deltas).

The bulk of the IETF's work is conducted on mailing lists. Multicast specific work is conducted on IP over ATM working group's list. To added to this list, send mail to ip-atm-request@matmos.hpl.hp.com. To send mail to the entire list, send mail to ip-atm@matmos.hpl.hp.com.

A1.1 Some Definitions.

Unicast, multicast, and broadcast may be defined in terms of the number of destinations reached by a single transmit operation of a single PDU by a source interface. In this contribution they are defined as follows:

Unicasting - the PDU reaches a single destination.

Multicasting - the PDU reaches a group of one or more destinations.

Broadcasting - the PDU reaches all destinations.

The astute will notice that both unicast and broadcast transmissions are special cases of multicast, with a groups of one or all destinations respectively.

(In practice the definition of broadcast is supplemented by some statement defining the scope of 'all destinations'. In LAN environments the physical size of a link level segment usually defines the scope of a broadcast. At the layer 3 level (e.g. IP) the scope of a broadcast is usually defined by the scope of a unicast subnet (however 'subnet' is defined)).

We are aware that with ATM networks it can be foolish to define 'all destinations' to cover every attached ATM interface. In practice the interfaces attached to an ATM network are usually already divided into logical or virtual layer 3 (sub)networks for unicast purposes, providing a reasonable basis on which to define the scope of a broadcast.

For the purpose of this contribution an abstract term will be borrowed from the current IETF work in this area [A1]:

Cluster - the set of ATM interfaces that are willing and able to participate in direct ATM connections to achieve multicasting and broadcasting of AAL_SDUs between themselves.

A1.2 The General Problem for MPOA.

A number of questions require answering before we move towards specifying a solution:

- How do we achieve efficient multicasting of AAL_SDUs around a Cluster (intra-cluster)?

- How do the solutions we look for here interact with our unicast solutions? (and should the unicast solution simply be a subset of the general multicast solution?)
- What minimal UNI functionality will MPOA expect to work with?

The last question has a large impact on the complexity of MPOA clients. However, we do not really have many choices. If we wish MPOA to have a solution useful to the present and growing base of installed ATM networks, we must support intra-cluster multicasting using UNI 3.1.

The second last question has an obvious technical answer - that we should build the MPOA unicast and broadcast support around a core, general multicast mechanism. However, it poses an equally obvious practical and political problem - most people are already further down the development path for unicast-only solutions.

A2. Intra-cluster multicasting using UNI 3.1

Two key areas need to be addressed - the data path, and group management.

The data path covers the actual mechanism(s) used to achieve the replication and redistribution of a source's AAL_SDUs to the destinations. Group management covers the mechanisms used to ensure the set of destinations can change over time, in response to the changing needs of the overlying Layer 3 protocol(s).

A2.1 Data path - methods of multicasting.

Given the definition of multicasting in section 1.1, the source must be able to execute a single transmit operation to multicast an AAL_SDU. Assume the source is an AAL User passing an AAL_SDU to a single instance of an AAL5 service. This implies that traffic exits the source interface and crosses the UNI on a single VC - multicasting occurs beyond the UNI.

UNI 3.1 provides the ability for a source interface to originate a point-to-multipoint VC, which is exactly what our model requires from the source's perspective. Using this basic service, two key models may be implemented - 'multicast VC meshes' and 'multicast servers'. In both cases the AAL_SDUs are passed transparently from source to destination(s), without interpretation of their contents at any point along their path.

The most fundamental approach, given the existence of point-to-multipoint VCs under UNI 3.1, is the multicast mesh, shown in Figure 1(a). This approach results in a data path from source to destination(s) where the actual multicasting occurs down at the cell level, within the ATM cloud's switch fabrics. Multipoint-to-multipoint communication is achieved when multiple senders establish their own data paths to the same set of leaf (receiving) nodes, giving rise to the name 'mesh'.

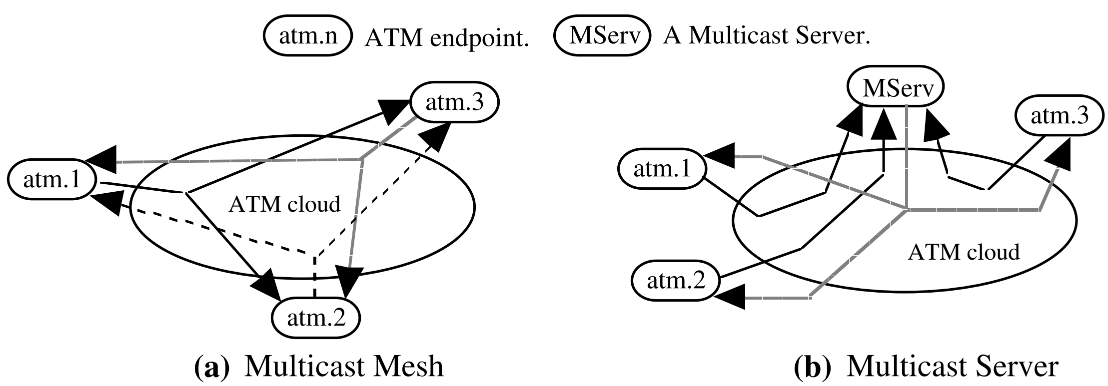


Figure 8 - ATM Level Multicasting - Distributed Meshes or Central Servers

The 'multicast server' model is shown in Figure 4(b). In this model the source(s) establish a point-to-multipoint VC with only a single leaf - the multicast server. The multicast server itself establishes and manages a point-to-multipoint VC out to the actual desired destinations. (Although a point-to-multipoint VC is not strictly necessary in figure 4(b), more complex architectures may use more than one multicast server to distribute the traffic for a given layer 3 multicast group. In this case each multicast server would be one leaf node on the source's outgoing VC).

AAL_SDUs sent to the multicast server are reassembled and then retransmitted back out to all endpoints that have been identified as receivers of this traffic. The leaf nodes of the multicast server's point-to-multipoint VC must be established prior to packet transmission, and the multicast server requires an external mechanism to identify them. (An alternative method is for the multicast server to explicitly retransmit packets on individual VCs between itself and group members. A benefit of this second approach is that the multicast server can ensure that sources do not receive copies of their own packets. However, this is significantly less efficient.)

Reassembly of incoming AAL_SDUs is required at a multicast server as AAL5 does not support cell level multiplexing of different AAL_SDUs on outgoing VCs.

A2.2 Group management and addressing.

Under UNI 3.1 a point-to-multipoint VC is established and managed by the root (source) interface. Only the root may add or drop leaf nodes on this VC, and it must explicitly identify the leaf nodes using their unicast ATM addresses. There is no concept of an ATM group address under UNI 3.1.

This is the core of an essential mismatch between UNI 3.1 and the requirements of protocols such as IP. The IP multicasting model is receiver controlled - leaf nodes decide when they'll become leaf nodes [A2]. UNI 3.1's model is source controlled - the data path only changes when the root issues an ADD_PARTY or DROP_PARTY.

This mismatch imposes a requirement that MPOA's solution has the ability to propagate the group join/leave intentions of Layer 3 entities back to the active source(s) for the group being joined or left. Further, this propagation mechanism must be timely.

Layer 3 data protocols that support the notion of multicasting usually do so through the notion of 'group addresses' - a destination address that indirectly identifies the set of layer 3 group members to which a packet must be sent. A node becomes a source as soon as it is required to transmit a packet with a layer 3 group address as its destination. At this point the source will need to ascertain what unicast ATM addresses correspond to the leaf nodes of the group identified by the specified layer 3 destination group address.

The MPOA solution must therefore provide a mechanism for a new source to find out who all a group's current leaf nodes are.

The MPOA multicast solution will need to support the mapping of arbitrary layer 3 addresses to sets of one or more {ATM.1, ATM.2, ..., ATM.n} addresses. Further, the addresses identified by ATM.n may specify either an NSAPA, E.164, or combined Called Party Number and Called Party Subaddress.

A2.3 VC meshes and Multicast Servers - a time and place for either.

Arguments over the relative merits of VC meshes and multicast servers have raged for some time. Ultimately the choice depends on the relative trade-offs a system architect must make between throughput, latency, congestion, and resource consumption. Even criteria such as latency can mean different things to different people - is it end to end packet time, or the time it takes for a group to settle after a membership change? Ultimately the decision depends on the characteristics of the applications generating the multicast traffic.

If we focused on the data path we might prefer the VC mesh because it lacks the obvious single congestion point of a multicast server. 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 multicast servers. The underlying ATM signaling system also has greater opportunity to ensure optimal branching points at switches along the multicast trees coming out of each source.

However, resource consumption will be higher. Every group member's ATM interface must terminate a VC per source (consuming space for state information, instance of an AAL service, and buffering in accordance with the vendors particular architecture). On the contrary, with a multicast server only 2 VCs (one out, one in) are required, independent of the number of senders. The allocation of VC related resources is also lower within the ATM cloud when using a multicast server. These points may be considered to have merit in environments where VCs across the UNI or within the ATM cloud are valuable (e.g. the ATM provider charges on a per VC basis), or AAL contexts are limited in the ATM interfaces of endpoints.

If we focus on the signaling load then multicast servers have the advantage when faced with dynamic sets of receivers. Every time the membership of a multicast group changes (a leaf node needs to be added or dropped), only a single point-to-multipoint VC needs to be modified when using a multicast server. This maps to a single ADD_PARTY or DROP_PARTY message across the UNI. When your multicast service is based on a mesh, an ADD or DROP action is triggered at the UNIs of every traffic source - the transient signaling load scales with the number of sources. This has obvious ramifications if you define latency as the time for a group to stabilize after change.

Finally, multicast servers introduce a 'reflected packet' problem. Sources that are also group members will get copies of their own packets straight back from the multicast server. The MPOA solution must ensure sufficient information is carried within each PDU to enable identification and removal of these reflected packets (this problem was faced and addressed in LANE by the use of a LEC_ID in the packet encapsulation - a more general, LLC/SNAP based solution, is being discussed within the IP-ATM working group [A3]).

The MPOA solution should allow system administrators to utilize either approach on a group by group basis.

A2.4 Broadcast as a special case of multicast.

Treat a layer 3 specific 'broadcast' address as referring to a group containing the entire membership of a Cluster. Allow the MPOA multicast address mapping service to return all members of a Cluster in response to a query for the members of the Layer 3 protocol's 'broadcast' group.

Layer 3 broadcasts are dangerous things to encourage. However, the transition of existing legacy applications will be expedited if support for this is included in MPOA. It is trivial to include it once the general multicast support exists.

A3. Summary of Demands Placed on MPOA's Multicast Solution.

- If we wish MPOA to have a solution useful to the present and growing base of installed ATM networks, we must support intra-cluster multicasting using UNI 3.1.
- We should consider building the MPOA broadcast support around a core, general multicast mechanism.
- The MPOA multicast solution will need to support the mapping of arbitrary layer 3 addresses to sets of one or more {ATM.1, ATM.2, ..., ATM.n} addresses.
- The MPOA multicast solution requires the ability to propagate the group join/leave intentions of Layer 3 entities back to the active source(s) for the group being joined or left.
- The MPOA multicast solution should allow system administrators to utilize either multicast servers or VC meshes on a group by group basis.

- The MPOA packet encapsulation mechanism must provide support for the detection and filtering of reflected packets when multicast servers are in use.

A4. Appendix A References

[A1] Grenville Armitage, “support for Multicast over UNI 3.1 based ATM Networks.”, INTERNET-DRAFT <draft-ietf-ipatm-ipmc-04.txt>, IP over ATM Working Group, February, 1995.

[A2] S. Deering, “Host Extensions for IP Multicasting”, RFC 1112, Stanford University, August 1989.

[A3] G. Armitage, “Issues surrounding a new encapsulation for IP over ATM”, INTERNET-DRAFT <draft-armitage-ipatm-encaps-00.txt>, IP over ATM Working Group, April 1995.

[A4] T. J. Smith, G. Armitage, “IP Broadcast over ATM networks”, INTERNET-DRAFT <draft-smith-ipatm-bcast-00.txt>, IP over ATM Working Group, April, 1995

[A5] G. Armitage, “Using the MARS to support IP unicast over ATM”, INTERNET-DRAFT <draft-armitage-ipatm-mars-unicast-00.txt>, IP over ATM Working Group, April, 1995.

For Further Reading:

http://gump.bellcore.com:8000/ ~gja/i-drafts.html

Appendix B - Resolution Message Contents vs. NHRP

B1. Resolution Request Contents

NHRP Field [3]	Associated MPOA Field	Comment(s)/Reason Added
Version	Version	P/O NHRP Fixed Header
Hop Count	Hop Count	P/O NHRP Fixed Header
Checksum	Checksum	Computed for the entire message - P/O NHRP Fixed Header
Type	Message Type	P/O NHRP Fixed Header
Mandatory Part Length	Mandatory Part Length	P/O NHRP Fixed Header
Q-Flag	Origin Type	Optional
S-Flag	Destination Type	Clear in request
A-Flag	Authority	Authoritative Indication - Always set for MPOA
P-Flag	Status	Clear in request
B-Flag	Stability	Clear in request
	EDFG Origin	Uses unused NHRP bits (NHRP flags field) set if origin is an EDFG
Protocol ID	Protocol ID	Based on ISO TR-9577:1990/3 - SNAP+ (total 6 bytes)
Request ID	Transaction ID	Request/response match
	Address Length	Place holder for non-fixed length addresses
Destination IP Address	Destination Inter-network Address	
	Address Length	Place holder for non-fixed length addresses
Source IP Address	Origin Inter-network Address	
	List Length	zero in request

B2. Resoluton Response Contents

NHRP Field [3]	Associated MPOA Field	Comment(s)/Reason Added
Version	Version	Optional in MPOA resolution response - P/O NHRP Fixed Header
Hop Count	Hop Count	P/O NHRP Fixed Header
Checksum	Checksum	Computed for the entire message -P/O NHRP Fixed Header
Type	Message Type	P/O NHRP Fixed Header
Mandatory Part Length	Mandatory Part Length	P/O NHRP Fixed Header
Q-Flag	Origin Type	Copied from request
S-Flag	Destination Type	Optional
A-Flag	Authority	Authoritative Indication - Always set for MPOA
P-Flag	Status	Non-zero value indicates
B-Flag	Stability	FailureMapping guarantee

	EDFG Origin	Uses unused NHRP bits (NHRP flags field) set if origin is an EDFG
Protocol ID	Protocol ID	Based on ISO TR-9577:1990/3 - SNAP+ (total 6 bytes)
Request ID	Transaction ID	Copied from request.
	Address Length	See corresponding request field.
Destination IP Address	Destination Inter-network Address	
	Address Length	See corresponding request field.
Source IP Address	Origin Inter-network Address	
	List Length	Number of MPOA Target entries in response.
	MPOA Target	One or more MPOA target entries.

B3. MPOA Target

NHRP Field [3]	Associated MPOA Field	Comment(s)/Reason Added
Next Hop IP Internetwork Address	Next Hop Inter-network Address	Address of RSFG local to the MPOA Target
Source Holding Time	Source Holding Time	Useful un-refreshed lifetime of this Response
Preference	Preference	Preference of this entry relative to other entries
Address Type	ATM Address Type	
NBMA Length	ATM Address Length	
NBMA Address	Termination ATM Address	
	AAL5 Max SDU size	Optional
	Destination Type	Clear if destination is a host
	QoS Available	Optional