Internet Engineering Task Force
INTERNET-DRAFT
draft-ietf-ipatm-framework-doc-05

R. G. Cole
D. H. Shur
AT&T Bell Laboratories
C. Villamizar
ANS

September 5, 1995

**IP over ATM: A Framework Document**

# Status of this Memo



# Abstract

The discussions of the IP over ATM working group over the last several years have produced a diverse set of proposals, some of which are no longer under active consideration. A categorization is provided for the purpose of focusing discussion on the various proposals for IP over ATM deemed of primary interest by the IP over ATM working group. The intent of this framework is to help clarify the differences between proposals and identify common features in order to promote convergence to a smaller and more mutually compatible set of standards. In summary, it is hoped that this document, in classifying ATM approaches and issues will help to focus the IP over ATM working group's direction.

# 1 Introduction

The IP over ATM Working Group of the Internet Engineering Task Force (IETF) is chartered to develop standards for routing and forwarding IP packets over ATM sub-networks. This document provides a classification/taxonomy of IP over ATM options and issues and then describes various proposals in these terms.

The remainder of this memorandum is organized as follows:

- Section 2 defines several terms relating to networking and internetworking.
- Section 3 discusses the parameters for a taxonomy of the different ATM models under discussion.
- Section 4 discusses the options for low level encapsulation.
- Section 5 discusses tradeoffs between connection oriented and connectionless approaches.
- Section 6 discusses the various means of providing direct connections across IP subnet boundaries.
- Section 7 discusses the proposal to extend IP routing to better accommodate direct connections across IP subnet boundaries.
- Section 8 identifies several prominent IP over ATM proposals that have been discussed within the IP over ATM Working Group and their relationship to the framework described in this document.
- Section 9 addresses the relationship between the documents developed in the IP over ATM and related working groups and the various models discussed.

## 2 Definitions and Terminology

We define several terms:

A Host or End System (ES): A host delivers/receives IP packets to/from other systems, but does not relay IP packets.

A Router or Intermediate System (IS): A router delivers/receives IP packets to/from other systems, and relays IP packets among systems.

IP Subnet: In an IP subnet, all members of the subnet are able to transmit packets to all other members of the subnet directly, without forwarding by intermediate entities.

Bridged IP Subnet: A bridged IP subnet is one in which two or more physically disjoint media are made to appear as a single IP subnet. There are two basic types of bridging, media access control (MAC) level, and proxy ARP (see section 6).

A Broadcast Subnet: A broadcast network supports an arbitrary number of hosts and routers and additionally is capable of transmitting a single IP packet to all of these systems.

A Multicast Capable Subnet: A multicast capable subnet supports a facility to send a packet which reaches a subset of the destinations on the subnet. Multicast setup may be sender initiated, or leaf initiated. ATM UNI 3.0 [4] and UNI 3.1 support only sender initiated while IP supports leaf initiated join. UNI 4.0 will support leaf initiated join.

A Non-Broadcast Multiple Access (NBMA) Subnet: An NBMA supports an arbitrary number of hosts and routers but does not support a convenient multi-destination connectionless transmission facility, as does a broadcast or multicast capable subnetwork.

An End-to-End path: An end-to-end path consists of two hosts which can communicate with one another over an arbitrary number of routers and subnets.

An internetwork: An internetwork (small "i") is the concatenation of networks, often of various different media and lower level encapsulations, to form an integrated larger network supporting communication between any of the hosts on any of the component networks. The Internet (big "I") is a specific well known global concatenation of (over 40,000 at the time of writing) component networks.

IP forwarding: IP forwarding is the process of receiving a packet and using a very low overhead decision process, determining how to forward the packet.

IP routing: IP routing is the exchange of information that takes place in order to have available the information necessary to make a correct IP forwarding decision.

IP address resolution: A quasi-static mapping exists between IP address on the local IP subnet and media address on the local subnet. This mapping is known as IP address resolution. An address resolution protocol (ARP) is a protocol supporting address resolution.

In order to support end-to-end connectivity, two techniques are used. One involves allowing direct connectivity across classic IP subnet boundaries supported by certain NBMA media, which includes ATM. The other involves IP routing and IP forwarding. In essence, the former technique is extending IP address resolution beyond the boundaries of the IP subnet, while the latter is interconnecting IP subnets.

Large internetworks, and in particular the Internet, are unlikely to be composed of a single media, or a star topology, with a single media at the center. Within a large network supporting a common media, typically any large NBMA such as ATM, IP routing and IP forwarding must always be accommodated if the internetwork is larger than the NBMA, particularly if there are multiple points of interconnection with the NBMA and/or redundant, diverse interconnections.

Routing information exchange in a very large internetwork can be quite dynamic due to the high probability that some network elements are changing state. The address resolution space consumption and resource consumption due to state change, or maintenance of state information is rarely a problem in classic IP subnets. It can become a problem in large bridged networks or in proposals that attempt to extend address resolution beyond the IP subnet. Scaling properties of address resolution and routing proposals, with respect to state information and state change, must be considered.

# 3 Parameters Common to IP Over ATM Proposals

In some discussion of IP over ATM distinctions have made between local area networks (LANs), and wide area networks (WANs) that do not necessarily hold. The distinction between a LAN, MAN and WAN is a matter of geographic dispersion. Geographic dispersion affects performance due to increased propagation delay.

LANs are used for network interconnections at the the major Internet traffic interconnect sites. Such LANs have multiple administrative authorities, currently exclusively support routers providing transit to multihomed internets, currently rely on PVCs and static address resolution, and rely heavily on IP routing. Such a configuration differs from the typical LANs used to interconnect computers in corporate or campus environments, and emphasizes the point that prior characterization of LANs do not necessarily hold. Similarly, WANs such as those under consideration by numerous large IP providers, do not conform to prior characterizations of ATM WANs in that they have a single administrative authority and a small number of nodes aggregating large flows of traffic onto single PVCs and rely on IP routers to avoid forming congestion bottlenecks within ATM.

The following characteristics of the IP over ATM internetwork may be independent of geographic dispersion (LAN, MAN, or WAN).

- The size of the IP over ATM internetwork (number of nodes).
- The size of ATM IP subnets (LIS) in the ATM Internetwork.
- Single IP subnet vs multiple IP subnet ATM internetworks.
- Single or multiple administrative authority.
- Presence of routers providing transit to multihomed internets.
- The presence or absence of dynamic address resolution.
- The presence or absence of an IP routing protocol.

IP over ATM should therefore be characterized by:

- Encapsulations below the IP level.
- Degree to which a connection oriented lower level is available and utilized.
- Type of address resolution at the IP subnet level (static or dynamic).
- Degree to which address resolution is extended beyond the IP subnet boundary.
- The type of routing (if any) supported above the IP level.

ATM-specific attributes of particular importance include:

- The different types of services provided by the ATM Adaptation Layers (AAL). These specify whether a connection oriented or a connectionless service is provided, the Quality-of-Service, the connection-mode, etc. The models discussed within this document assume connection-oriented, best effort services over AAL Type 5.
- The type of virtual circuits used, i.e., PVCs versus SVCs. The PVC environment requires the use of either static tables for ATM-to-IP address mapping or the use of inverse ARP, while the SVC environment requires ARP functionality to be provided.
- The type of support for multicast services. If point-to-point services only are available, then a server for IP multicast is required. If point-to-multipoint services are available, then IP multicast can be supported via meshes of point-to-multipoint connections (although use of a server may be necessary due to limits on the number of multipoint VCs able to be supported or to maintain the leaf initiated join semantics).

- The presence of logical link identifiers (VPI/VCIs) and the various information element (IE) encodings within the ATM SVC signaling specification, i.e., the ATM Forum UNI version 3.1. This allows a VC originator to specify a range of "layer" entities as the destination "AAL User". The AAL specifications do not prohibit any particular "layer X" from attaching directly to a local AAL service. Taken together these points imply a range of methods for encapsulation of upper layer protocols over ATM. For example, while LLC/SNAP encapsulation is one approach (the default), it is also possible to bind virtual circuits to higher level entities in the TCP/IP protocol stack. Some examples of the latter are single VC per protocol binding, TULIP, and TUNIC, discussed further in Section 4.
- The number and type of ATM administrative domains/networks, and type of addressing used within an administrative domain/network. In particular, in the single domain/network case, all attached systems may be safely assumed to be using a single common addressing format, while in the multiple domain case, attached stations may not all be using the same common format, with corresponding implications on address resolution. (See Appendix A for a discussion of some of the issues that arise when multiple ATM address formats are used in the same logical IP subnet (LIS).) Also security/authentication is much more of a concern in the multiple domain case.

IP over ATM proposals do not universally accept that IP routing over an ATM network is required. Certain proposals rely on the following assumptions:

- The widespread deployment of ATM within premises-based networks, private wide-area networks and public networks, and
- The definition of interfaces, signaling and routing protocols among private ATM networks.

The above assumptions amount to ubiquitous deployment of a seamless ATM fabric which serves as the hub of a star topology around which all other media is attached. There has been a great deal of discussion over when, if ever, this will be a realistic assumption for very large internetworks, such as the Internet. Advocates of such approaches point out that even if these are not relevant to very large internetworks such as the Internet, there may be a place for such models in smaller internetworks, such as corporate networks.

The NHRP protocol (Section 8.2), not necessarily specific to ATM, would be particularly appropriate for the case of ubiquitous ATM deployment. NHRP supports the establishment of direct connections across IP subnets in the ATM domain. The use of NHRP does not require ubiquitous ATM deployment, but currently imposes topology constraints to avoid routing loops (see Section 7). Section 8.2 describes NHRP in greater detail.

The Peer Model assumes that internetwork layer addresses can be mapped onto ATM addresses and vice versa, and that reachability information between ATM routing and internetwork layer routing can be exchanged. This approach has limited applicability unless ubiquitous deployment of ATM holds. The peer model is described in Section 8.4.

The Integrated Model proposes a routing solution supporting an exchange of routing information between ATM routing and higher level routing. This provides timely external routing information within the ATM routing and provides transit of external routing information through the ATM routing between external routing domains. Such proposals may better support a possibly lengthy transition during which assumptions of ubiquitous ATM access do not hold. The Integrated Model is described in Section 8.5.

The Multiprotocol over ATM (MPOA) Sub-Working Group was formed by the ATM Forum to provide multiprotocol support over ATM. The MPOA effort is at an early stage at the time of this writing. An MPOA baseline document has been drafted, which provides terminology for further discussion of the architecture.

# 4 Encapsulations and Lower Layer Identification

Data encapsulation, and the identification of VC endpoints, constitute two important issues that are somewhat orthogonal to the issues of network topology and routing. The relationship between these two issues is also a potential sources of confusion. In conventional LAN technologies the 'encapsulation' wrapped around a packet of data typically defines the (de)multiplexing path within source and destination nodes (e.g. the Ethertype field of an Ethernet packet). Choice of the protocol endpoint within the packet's destination node is essentially carried 'in-band'.

As the multiplexing is pushed towards ATM and away from LLC/SNAP mechanism, a greater burden will be placed upon the call setup and teardown capacity of the ATM network. This may result in some questions being raised regarding the scalability of these lower level multiplexing options.

With the ATM Forum UNI version 3.1 service the choice of endpoint within a destination node is made 'out of band' - during the Call Setup phase. This is quite independent of any in-band encapsulation mechanisms that may be in use. The B-LLI Information Element allows Layer 2 or Layer 3 entities to be specified as a VC's endpoint. When faced with an incoming SETUP message the Called Party will search locally for an AAL User that claims to provide the service of the layer specified in the B-LLI. If one is found then the VC will be accepted (assuming other conditions such as QoS requirements are also met).

An obvious approach for IP environments is to simply specify the Internet Protocol layer as the VCs endpoint, and place IP packets into AAL–SDUs for transmission. This is termed 'VC multiplexing' or 'Null Encapsulation', because it involves terminating a VC (through an AAL instance) directly on a layer 3 endpoint. However, this approach has limitations in environments that need to support multiple layer 3 protocols between the same two ATM level endpoints. Each pair of layer 3 protocol entities that wish to exchange packets require their own VC.

RFC–1483 [6] notes that VC multiplexing is possible, but focuses on describing an alternative termed 'LLC/SNAP Encapsulation'. This allows any set of protocols that may be uniquely identified by an LLC/SNAP header to be multiplexed onto a single VC. Figure 1 shows how this works for IP packets - the first 3 bytes indicate that the payload is a Routed Non-ISO PDU, and the Organizationally Unique Identifier (OUI) of 0x00-00-00 indicates that the Protocol Identifier (PID) is derived from the EtherType associated with IP packets (0x800). ARP packets are multiplexed onto a VC by using a PID of 0x806 instead of 0x800.

Whatever layer terminates a VC carrying LLC/SNAP encapsulated traffic must know how to parse the AAL–SDUs in order to retrieve the packets. The recently approved signalling standards for IP over ATM are more explicit, noting that the default SETUP message used to establish IP over ATM VCs must carry a B-LLI specifying an ISO 8802/2 Layer 2 (LLC) entity as each VCs endpoint. More

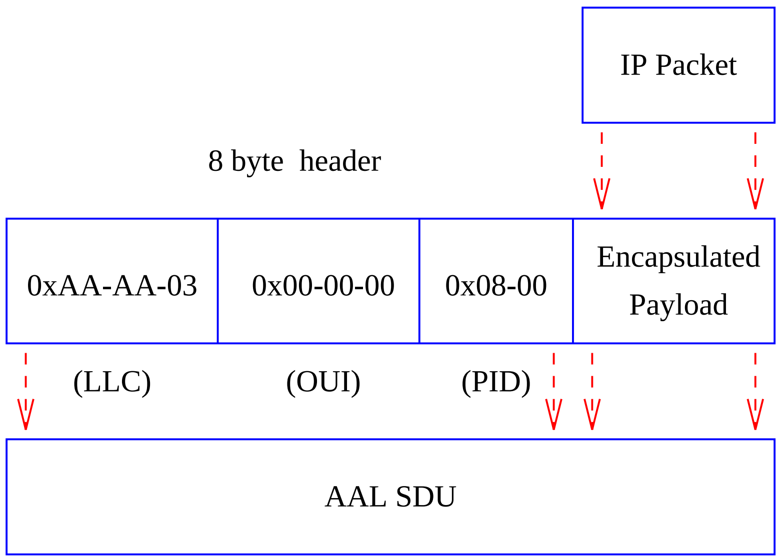


Figure 1: AAL5 PDU is an encapsulated IP packet

significantly, there is no information carried within the SETUP message about the identity of the layer 3 protocol that originated the request - until the packets begin arriving the terminating LLC entity cannot know which one or more higher layers are packet destinations.

Taken together, this means that hosts require a protocol entity to register with the host's local UNI 3.1 management layer as being an LLC entity, and this same entity must know how to handle and generate LLC/SNAP encapsulated packets. The LLC entity will also require mechanisms for attaching to higher layer protocols such as IP and ARP. Figure 2 attempts to show this, and also highlights the fact that such an LLC entity might support many more than just IP and ARP. In fact the combination of RFC 1483 LLC/SNAP encapsulation, LLC entities terminating VCs, and suitable choice of LLC/SNAP values, can go a long way towards providing an integrated approach to building multiprotocol networks over ATM.

The processes of actually establishing AAL Users, and identifying them to the local UNI 3.1 management layers, are still undefined and are likely to be very dependent on operating system environments.

Two encapsulations have been discussed within the IP over ATM working group which differ from those given in RFC–1483 [6]. These have the characteristic of largely or totally eliminating IP header overhead. These models were discussed in the July 1993 IETF meeting in Amsterdam, but have not been fully defined by the working group.

Given a single hop configuration between Sub-ES, both Sub-ES assume that following name resolution, address resolution, and SVC signaling, an implicit binding is established between entities in the two Sub-ESes. In this case full IP headers (and in particular source and destination addresses) are not required in each data packet.

- The first model is "TCP and UDP over Lightweight IP" (TULIP) in which only the IP protocol field is carried in each packet, everything else being bound at call set-up time. In this case the implicit binding is between the IP entities in each Sub-ES. Since there is no further routing problem once the binding is established, since AAL5 can indicate packet size, since fragmen-

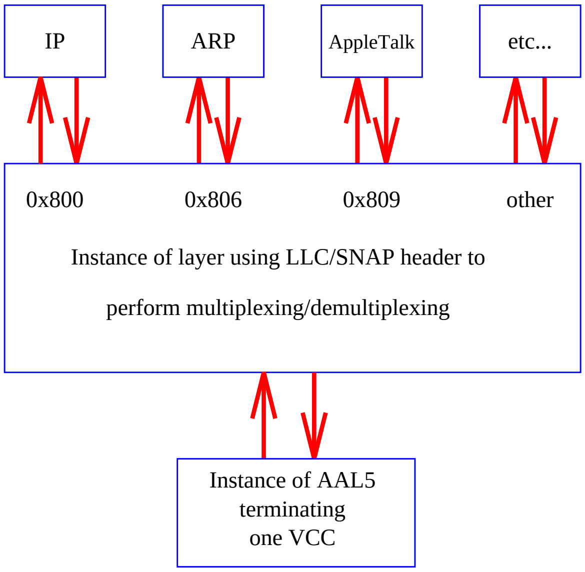


Figure 2: LLC/SNAP encapsulation allows more than just IP or ARP per VC.

tation cannot occur, and since ATM signaling will handle exception conditions, the absence of all other IP header fields and of ICMP should not be an issue. Entry to TULIP mode would occur as the last stage in SVC signaling, by a simple extension to the encapsulation negotiation described in RFC–1755 [10].

TULIP changes nothing in the abstract architecture of the IP model, since each subnet ES still has an IP address which is resolved to an ATM address. It simply uses the point-to-point property of VCs to allow the elimination of some per-packet overhead. The use of TULIP could in principle be negotiated on a per-SVC basis or configured on a per-PVC basis.

- The second model is "TCP and UDP over a Nonexistent IP Connection" (TUNIC). In this case no network-layer information is carried in each packet, everything being bound at virtual circuit set-up time. The implicit binding is between two applications using either TCP or UDP directly over AAL5 on a dedicated VC. If this can be achieved, the IP protocol field has no useful dynamic function. However, in order to achieve binding between two applications, the use of a well-known port number in classical IP or in TULIP mode may be necessary during call set-up. This is a subject for further study and would require significant extensions to the use of SVC signaling described in RFC–1755 [10].

TULIP/TUNIC can be presented as being on one end of a continuum opposite the SNAP/LLC encapsulation, with various forms of null encapsulation somewhere in the middle. The continuum is simply a matter of how much is moved from in-stream demultiplexing to call setup demultiplexing. The various encapsulation types are presented in Table 1.

Encapsulations such as TULIP and TUNIC make assumptions with regard to the desirability to support connection oriented flow. The tradeoffs between connection oriented and connectionless are discussed in Section 5.

| Encapsulation | In setup message | Demultiplexing |
|---|---|---|
| SNAP/LLC | nothing | source and destination address, protocol family, protocol, ports |
| NULL encaps | protocol family | source and destination address, protocol, ports |
| TULIP | source and destination address, protocol family | protocol, ports |
| TUNIC - A | source and destination address, protocol family protocol | ports |
| TUNIC - B | source and destination address, protocol family protocol, ports | nothing |

Table 1: Summary of Encapsulation Types

# 5 Connection Oriented and Connectionless Tradeoffs

The connection oriented and connectionless approaches each offer advantages and disadvantages. In the past, strong advocates of pure connection oriented and pure connectionless architectures have argued intensely. IP over ATM does not need to be purely connectionless or purely connection oriented.

IP over ATM provides a connection oriented architecture (ATM with AAL5) underlying a connectionless architecture (IP), on top of which much of the traffic is supported by a reliable end-to-end connection oriented protocol (TCP). A fundamental question is to what degree is it beneficial to map different flows above IP into separate connections below IP. There is a broad spectrum of opinion on this.

As stated in section 4, at one end of the spectrum, IP would remain highly connectionless and set up single VCs between routers which are adjacent on an IP subnet and for which there was active traffic flow. All traffic between the such routers would be multiplexed on a single ATM VC. At the other end of the spectrum, a separate ATM VC would be created for each identifiable flow. For every unique TCP or UDP address and port pair encountered a new VC would be required. Part of the intensity of early arguments has been over failure to recognize that there is a middle ground.

ATM offers QoS and traffic management capabilities that are well suited for certain types of services. It may be advantageous to use separate ATM VC for such services. Other IP services such as DNS, are ill suited for connection oriented delivery, due to their normal very short duration (typically one packet in each direction). Short duration transactions, even many using TCP, may also be poorly suited for a connection oriented model due to setup and state overhead. ATM QoS and traffic management capabilities may be poorly suited for elastic traffic.

| APPLICATION | Pure Connection Oriented Approach |
|---|---|
| General | Always set up a VC |
| Short Duration UDP (DNS) | Set up a VC. Either hold the packet during VC setup or drop it and await a retransmission. Teardown on a timer basis. |
| Short Duration TCP (SMTP) | Set up a VC. Either hold packet(s) during VC setup or drop them and await retransmission. Teardown on detection of FIN-ACK or on a timer basis. |
| Elastic (TCP) Bulk Transfer | Set up a VC same as above. No clear method to set QoS parameters has emerged. |
| Real Time (audio, video) | Set up a VC. QoS parameters are assumed to precede traffic in RSVP or be carried in some form within the traffic itself. |

Table 2: Connection Oriented vs. Connectionless - a) a pure connection oriented approach

| APPLICATION | Middle Ground |
|---|---|
| General | Use RSVP or other indication which clearly indicate a VC is needed and what QoS parameters are appropriate. |
| Short Duration UDP (DNS) | Forward hop by hop. RSVP is unlikely to precede this type of traffic. |
| Short Duration TCP (SMTP) | Forward hop by hop unless RSVP indicates otherwise. RSVP is unlikely to precede this type of traffic. |
| Elastic (TCP) Bulk Transfer | By default hop by hop forwarding is used. However, RSVP information, local configuration about TCP port number usage, or a locally implemented method for passing QoS information from the application to the IP/ATM driver may allow/suggest the establishment of direct VCs. |
| Real Time (audio, video) | Forward hop by hop unless RSVP indicates otherwise. RSVP will indicate QoS requirements. It is assumed RSVP will generally be used for this case. A local decision can be made as to whether the QoS is better served by a separate VC. |

Table 3: Connection Oriented vs. Connectionless - b) a middle ground approach

| APPLICATION | Pure Connectionless Approach |
|---|---|
| General | Always forward hop by hop. Use queueing algorithms implemented at the IP layer to support reservations such as those specified by RSVP. |
| Short Duration UDP (DNS) | Forward hop by hop. |
| Short Duration TCP (SMTP) | Forward hop by hop. |
| Elastic (TCP) Bulk Transfer | Forward hop by hop. Assume ability of TCP to share bandwidth (within a VBR VC) works as well or better than ATM traffic management. |
| Real Time (audio, video) | Forward hop by hop. Assume that queueing algorithms at the IP level can be designed to work with sufficiently good performance (e.g., due to support for predictive reservation). |

Table 4: Connection Oriented vs. Connectionless - c) a pure connectionless approach

Work in progress is addressing how QoS requirements might be expressed and how the local decisions might be made as to whether those requirements are best and/or most cost effectively accomplished using ATM or IP capabilities. Table 2, Table 3, and Table 4 describe typical treatment of various types of traffic using a pure connection oriented approach, middle ground approach, and pure connectionless approach.

The above qualitative description of connection oriented vs connectionless service serve only as examples to illustrate differing approaches. Work in the area of an integrated service model, QoS and resource reservation are related to but outside the scope of the IP over ATM Work Group. This work falls under the Integrated Services Work Group (int-serv) and Reservation Protocol Work Group (rsvp), and will ultimately determine *when* direct connections will be established. The IP over ATM Work Group can make more rapid progress if concentrating solely on *how* direct connections are established.

# 6 Crossing IP Subnet Boundaries

A single IP subnet will not scale well to a large size. Techniques which extend the size of an IP subnet in other media include MAC layer bridging, and proxy ARP bridging.

MAC layer bridging alone does not scale well. Protocols such as ARP rely on the media broadcast to exchange address resolution information. Most bridges improve scaling characteristics by capturing ARP packets and retaining the content, and distributing the information among bridging peers.

The ARP information gathered from ARP replies is broadcast only where explicit ARP requests are made. This technique is known as proxy ARP.

Proxy ARP bridging improves scaling by reducing broadcast traffic, but still suffers scaling problems. If the bridged IP subnet is part of a larger internetwork, a routing protocol is required to indicate what destinations are beyond the IP subnet unless a statically configured default route is used. A default route is only applicable to a very simple topology with respect to the larger internet and creates a single point of failure. Because internets of enormous size create scaling problems for routing protocols, the component networks of such large internets are often partitioned into areas, autonomous systems or routing domains, and routing confederacies.

The scaling limits of the simple IP subnet require a large network to be partitioned into smaller IP subnets. For NBMA media like ATM, there are advantages to creating direct connections across the entire underlying NBMA network. This leads to the need to create direct connections across IP subnet boundaries.

For example, figure 3 shows an end-to-end configuration consisting of four components, three of which are ATM technology based, while the fourth is a standard IP subnet based on non-ATM technology. End-systems (either hosts or routers) attached to the ATM-based networks may communicate either using the Classical IP model or directly via ATM (subject to policy constraints). Such nodes may communicate directly at the IP level without necessarily needing an intermediate router, even if end-systems do not share a common IP-level network prefix. Communication with end-systems on the non-ATM-based Classical IP subnet takes place via a router, following the Classical IP model (see Section 8.1 below).

Many of the problems and issues associated with creating such direct connections across subnet boundaries were originally being addressed in the IETF's IPLPDN working group and the IP over ATM working group. This area is now being addressed in the Routing over Large Clouds working group. Examples of work performed in the IPLPDN working group include short-cut routing (proposed by P. Tsuchiya) and directed ARP RFC–1433 [5] over SMDS networks. The ROLC working group has produced the distributed ARP server architectures and the NBMA Address Resolution Protocol (NARP) [7]. The Next Hop Resolution Protocol (NHRP) is still work in progress, though the ROLC WG is considering advancing the current draft. Questions/issues specifically related to defining a capability to cross IP subnet boundaries include:

- How can routing be optimized across multiple logical IP subnets over both a common ATM based and a non-ATM based infrastructure. For example, in Figure 3, there are two gateways/routers between the non-ATM subnet and the ATM subnets. The optimal path from end-systems on any ATM-based subnet to the non ATM-based subnet is a function of the routing state information of the two routers.
- How to incorporate policy routing constraints.
- What is the proper coupling between routing and address resolution particularly with respect to off-subnet communication.
- What are the local procedures to be followed by hosts and routers.
- Routing between end systems not sharing a common IP-level (or L3) network prefix, but able to be directly connected at the NBMA media level.
- Defining the details for an efficient address resolution architecture including defining the procedures to be followed by clients and servers (see RFC–1433 [5], RFC–1735 [7] and NHRP.

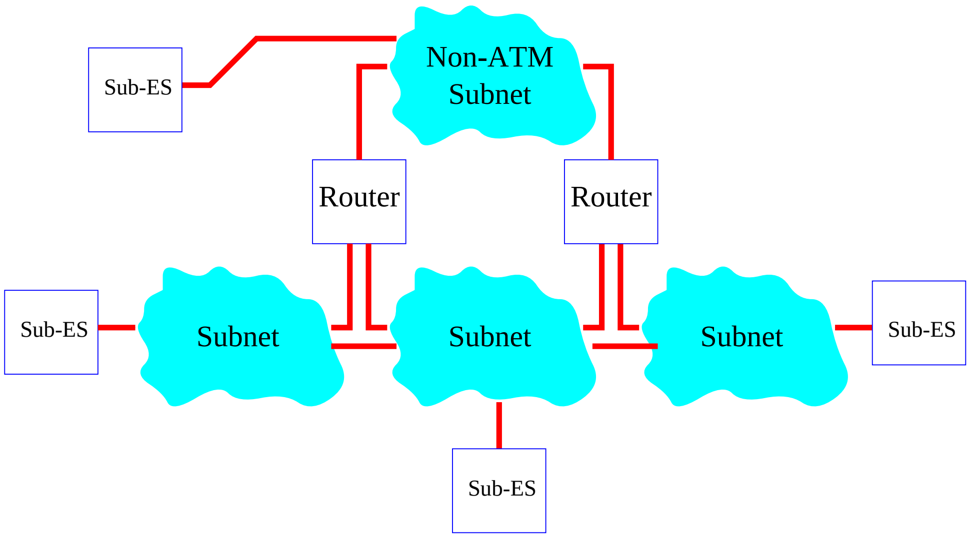


Figure 3: A configuration with both ATM-based and non-ATM based subnets.

- How to identify the need for and accommodate special purpose SVCs for control or routing and high bandwidth data transfers.

For ATM (unlike other NBMA media), an additional complexity in supporting IP routing over these ATM internets lies in the multiplicity of address formats in UNI 3.0 [4]. NSAP modeled address formats only are supported on "private ATM" networks, while either 1) E.164 only, 2) NSAP modeled formats only, or 3) both are supported on "public ATM" networks. Further, while both the E.164 and NSAP modeled address formats are to be considered as network points of attachment, it seems that E.164 only networks are to be considered as subordinate to "private networks", in some sense. This leads to some confusion in defining an ARP mechanism in supporting all combinations of end-to-end scenarios (refer to the discussion in Appendix A on the possible scenarios to be supported by ARP).

# 7 Extensions to IP Routing

RFC–1620 [3] describes the problems and issues associated with direct connections across IP subnet boundaries in greater detail, as well as possible solution approaches. The ROLC WG has identified persistent routing loop problems that can occur if protocols which lose information critical to path vector routing protocol loop suppression are used to accomplish direct connections across IP subnet boundaries. The problems may arise when a destination network which is not on the NBMA network is reachable via different routers attached to the NBMA network. This problem occurs with proposals that attempt to carry reachability information, but do not carry full path attributes (for path vector routing) needed for inter-AS path suppression, or full metrics (for distance vector or link state routing even if path vector routing is not used) for intra-AS routing.

There are many potential scenarios for routing loops. An example is given in Figure 4. It is possible to produce a simpler example where a loop can form. The example in Figure 4 illustrates a loop

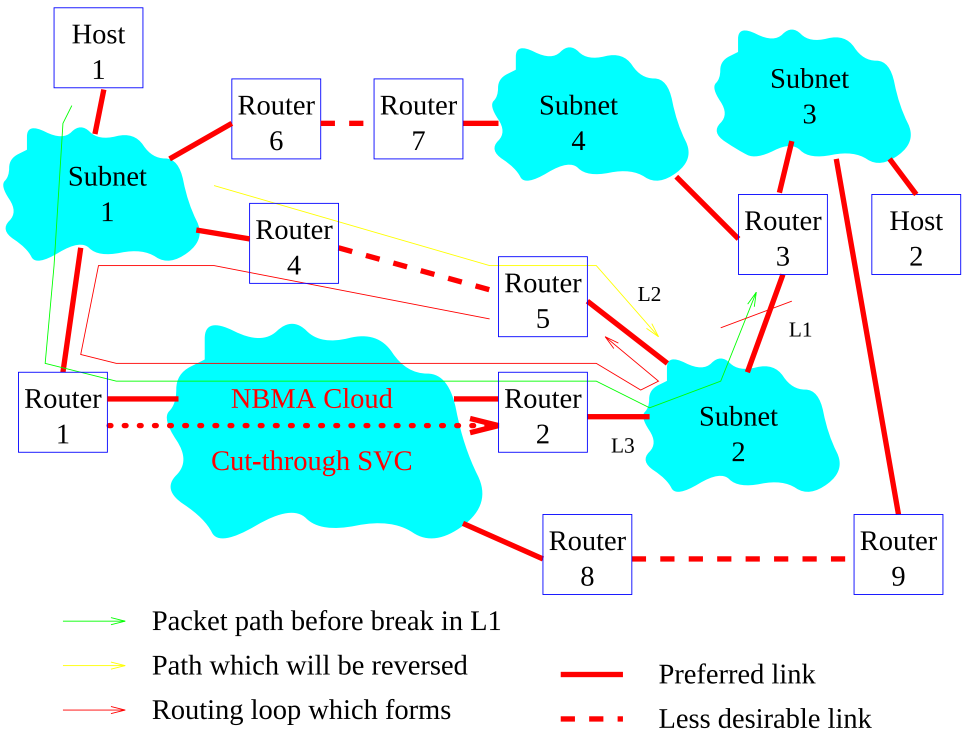


Figure 4: A Routing Loop Due to Lost PV Routing Attributes.

which will persist even if the protocol on the NBMA supports redirects or can invalidate any route which changes in any way, but does not support the communication of full metrics or path attributes.

In the example in Figure 4, Host 1 is sending traffic toward Host 2. In practice, host routes would not be used, so the destination for the purpose of routing would be Subnet 3. The traffic travels by way of Router 1 which establishes a "cut-through" SVC to the NBMA next-hop, shown here as Router 2. Router 2 forwards traffic destined for Subnet 3 through Subnet 2 to Router 3. Traffic from Host 1 would then reach Host 2.

Router 1's cut-through routing implementation caches an association between Host 2's IP address (or more likely all of Subnet 3) and Router 2's NBMA address. While the cut-through SVC is still up, Link 1 fails. Router 5 loses it's preferred route through Router 3 and must direct traffic in the other direction. Router 2 loses a route through Router 3, but picks up an alternate route through Router 5. Router 1 is still directing traffic toward Router 2 and advertising a means of reaching Subnet 3 to Subnet 1. Router 5 and Router 2 will see a route, creating a loop.

This loop would not form if path information normally carried by interdomain routing protocols such as BGP and IDRP were retained across the NBMA. Router 2 would reject the initial route from Router 5 due to the path information. When Router 2 declares the route to Subnet 3 unreachable, Router 1 withdraws the route from routing at Subnet 1, leaving the route through Router 4, which would then reach Router 5, and would reach Router 2 through both Router 1 and Router 5. Similarly, a link state protocol would not form such a loop.

Two proposals for breaking this form of routing loop have been discussed. Redirect in this example would have no effect, since Router 2 still has a route, just has different path attributes. A second

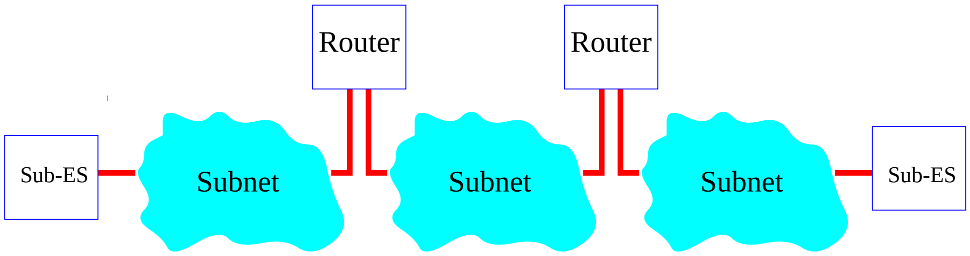


Figure 5: The Classical IP model as a concatenation of three separate ATM IP subnets.

proposal is that is that when a route changes in any way, the advertising NBMA cut-through router invalidates the advertisement for some time period. This is similar to the notion of Poison Reverse in distance vector routing protocols. In this example, Router 2 would eventually readvertise a route since a route through Router 6 exists. When Router 1 discovers this route, it will advertise it to Subnet 1 and form the loop. Without path information, Router 1 cannot distinguish between a loop and restoration of normal service through the link L1.

The loop in Figure 4 can be prevented by configuring Router 4 or Router 5 to refuse to use the reverse path. This would break backup connectivity through Router 8 if L1 and L3 failed. The loop can also be broken by configuring Router 2 to refuse to use the path through Router 5 unless it could not reach the NBMA. Special configuration of Router 2 would work as long as Router 2 was not distanced from Router 3 and Router 5 by additional subnets such that it could not determine which path was in use. If Subnet 1 is in a different AS or RD than Subnet 2 or Subnet 4, then the decision at Router 2 could be based on path information.

In order for loops to be prevented by special configuration at the NBMA border router, that router would need to know all paths that could lead back to the NBMA. The same argument that special configuration could overcome loss of path information was posed in favor of retaining the use of the EGP protocol defined in the now historic RFC–904 [11]. This turned out to be unmanageable, with routing problems occurring when topology was changed elsewhere.

# 8 IP Over ATM Proposals

## 8.1 The Classical IP Model

The Classical IP Model was suggested at the Spring 1993 IETF meeting [8] and retains the classical IP subnet architecture. This model simply consists of cascading instances of IP subnets with IP-level (or L3) routers at IP subnet borders. An example realization of this model consists of a concatenation of three IP subnets. This is shown in Figure 5. Forwarding IP packets over this Classical IP model is straight forward using already well established routing techniques and protocols.

SVC-based ATM IP subnets are simplified in that they:

- limit the number of hosts which must be directly connected at any given time to those that may actually exchange traffic.

- The ATM network is capable of setting up connections between any pair of hosts. Consistent with the standard IP routing algorithm [2] connectivity to the "outside" world is achieved only through a router, which may provide firewall functionality if so desired.
- The IP subnet supports an efficient mechanism for address resolution.

Issues addressed by the IP Over ATM Working Group, and some of the resolutions, for this model are:

- Methods of encapsulation and multiplexing. This issue is addressed in RFC–1483 [6], in which two methods of encapsulation are defined, an LLC/SNAP and a per-VC multiplexing option.
- The definition of an address resolution server (defined in RFC–1577).
- Defining the default MTU size. This issue is addressed in RFC–1626 [1] which proposes the use of the MTU discovery protocol (RFC–1191 [9]).
- Support for IP multicasting. In the summer of 1994, work began on the issue of supporting IP multicasting over the SVC LATM model. The proposal for IP multicasting is currently defined by a set of IP over ATM WG Internet Drafts, referred to collectively as the IPMC drafts. In order to support IP multicasting the ATM subnet must either support point-to- multipoint SVCs, or multicast servers, or both.
- Defining interim SVC parameters, such as QoS parameters and time-out values.
- Signaling and negotiations of parameters such as MTU size and method of encapsulation. RFC–1755 [10] describes an implementation agreement for routers signaling the ATM network to establish SVCs initially based upon the ATM Forum's UNI version 3.0 specification [4], and eventually to be based upon the ATM Forum's UNI version 3.1 and later specifications. Topics addressed in RFC–1755 include (but are not limited to) VC management procedures, e.g., when to time-out SVCs, QOS parameters, service classes, explicit setup message formats for various encapsulation methods, Sub-ES to Sub-ES negotiations, etc.

RFC-1577 is also applicable to PVC-based subnets. Full mesh PVC connectivity is required.

For more information see RFC–1577 [8].

## 8.2 The ROLC NHRP Model

The Next Hop Resolution Protocol (NHRP), currently a draft defined by the Routing Over Large Clouds Working Group (ROLC), performs address resolution to accomplish direct connections across IP subnet boundaries. NHRP can supplement RFC–1577 ARP. There has been recent discussion of replacing RFC–1577 ARP with NHRP. NHRP can also perform a proxy address resolution to provide the address of the border router serving a destination off of the NBMA which is only served by a single router on the NBMA. NHRP as currently defined cannot be used in this way to support addresses learned from routers for which the same destinations may be heard at other routers, without the risk of creating persistent routing loops.

## 8.3 Ohta's "Conventional" Model

The conventional model is layered and has the same network layer architecture as the standard IP model, hence it retains the IP subnet architecture. At the link layer, all types of ATM subnets are allowed. The basic difference between the Classical and the "Conventional" model is whether it is assumed that a router can relay IP packets cell by cell or not. Link layers are connected by routers but, when the two link layers are ATM, the routers are capable of forwarding cell-by-cell. This could provide a latency advantage depending on whether cell interleaving from multiple IP packets is allowed or not (e.g., an ATM AAL such as AAL3/4 would allow for cell interleaving, while AAL5 would not). This cell forwarding is accomplished through a higher level mapping, above the ATM VCI layer.

## 8.4 The Peer Model

The Peer Model places IP routers/gateways on an addressing peer basis with corresponding entities in an ATM cloud (where the ATM cloud may consist of a set of ATM networks, inter-connected via UNI or P-NNI interfaces). ATM network entities and the attached IP hosts or routers exchange call routing information on a peer basis by algorithmically mapping IP addressing into the NSAP space. Within the ATM cloud, ATM network level addressing (NSAP-style), call routing and packet formats are used. The Peer Model is being discussed in the Multiprotocol SWG of the ATM Forum.

In the Peer Model no provision is made for selection of primary path and use of alternate paths in the event of primary path failure in reaching multihomed non-ATM destinations. This will limit the topologies for which the peer model alone is applicable to only those topologies in which non-ATM networks are singly homed, or where loss of backup connectivity is not an issue. The Peer Model may be used to avoid the need for an address resolution protocol and in a proxy-ARP mode for stub networks, in conjunction with other mechanisms suitable to handle multihomed destinations.

During the discussions of the IP over ATM working group, it was felt that the problems with the end-to-end peer model were much harder than any other model, and had more unresolved technical issues. While encouraging interested individuals/companies to research this area, it was not an initial priority of the working group to address these issues. The ATM Forum Network Layer Multiprotocol Working Group is discussing this model.

## 8.5 The PNNI and the Integrated Models

The Integrated model (proposed and under study within the Multiprotocol group of ATM Forum) considers a single routing protocol to be used for both IP and for ATM. A single routing information exchange is used to distribute topological information. The routing computation used to calculate routes for IP will take into account the topology, including link and node characteristics, of both the IP and ATM networks and calculates an optimal route for IP packets over the combined topology.

The PNNI is a hierarchical link state routing protocol with multiple link metrics providing various available QoS parameters given current loading. Call route selection takes into account QoS requirements. Hysteresis is built into link metric readvertisements in order to avoid computational overload and topological hierarchy serves to subdivide and summarize complex topologies, helping to bound computational requirements.

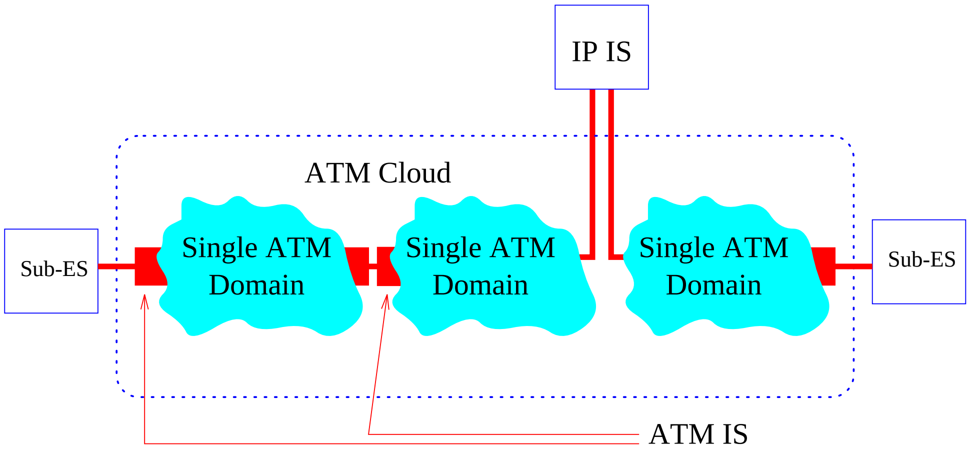


Figure 6: The ATM transition model assuming the presence of gateways or routers between the ATM networks and the ATM peer networks.

Integrated Routing is a proposal to use PNNI routing as an IP routing protocol. There are several sets of technical issues that need to be addressed, including the interaction of multiple routing protocols, adaptation of PNNI to broadcast media, support for NHRP, and others. These are being investigated. However, the ATM Forum MPOA group is not currently performing this investigation. Concerned individuals are, with an expectation of bringing the work to the ATM Forum and the IETF.

PNNI has provisions for carrying uninterpreted information. While not yet defined, a compatible extension of the base PNNI could be used to carry external routing attributes and avoid the routing loop problems described in Section 7.

## 8.6 Transition Models

Finally, it is useful to consider transition models, lying somewhere between the Classical IP Models and the Peer Models. Some possible architectures for transition models have been suggested by Fong Liaw. Others are possible, for example Figure 6 showing a Classical IP transition model which assumes the presence of gateways between ATM networks and ATM Peer networks.

Some of the models described in the prior sections, most notably the Integrated Model, anticipate the need for mixed environment with complex routing topologies. These inherently support transition (possibly with an indefinite transition period). Models which provide no transition support are primarily of interest to new deployments which make exclusive, or near exclusive use of ATM or deployments capable of wholesale replacement of existing networks or willing to retain only non-ATM stub networks.

For some models, most notably the Peer Model, the ability to attach to a large non-ATM or mixed internetwork is infeasible without routing support at a higher level, or at best may pose interconnection topology constraints (for example: single point of attachment and a static default route). If a particular model requires routing support at a higher level a large deployment will need to be subdivided to provide scalability at the higher level, which for some models degenerates back to the Classical model.

# 9 Application of the Working Group's and Related Documents

The IP Over ATM Working Group has generated several Internet-Drafts and RFCs. This section identifies the relationship of these and other related documents to the various IP Over ATM Models identified in this document. The Drafts and RFCs produced to date are the following references, RFC–1483 [6], RFC–1577 [8], RFC–1626 [1], RFC–1755 [10] and the IPMC drafts. The ROLC WG has produced the NHRP draft. Table 5 gives a summary of these documents and their relationship to the various IP Over ATM Models.

## Acknowledgments:

This draft is the direct result of the numerous discussions of the IP over ATM Working Group of the Internet Engineering Task Force. The authors also had the benefit of several private discussions with H. Nguyen of AT&T Bell Laboratories. Brian Carpenter of CERN was kind enough to contribute the TULIP and TUNIC sections to this draft. Grenville Armitage of Bellcore was kind enough to contribute the sections on VC binding, encapsulations and the use of B-LLI information elements to signal such bindings. The text of Appendix A was pirated liberally from Anthony Alles' of Cisco posting on the IP over ATM discussion list (and modified at the authors' discretion). M. Ohta provided a description of the Conventional Model (again which the authors modified at their discretion). This draft also has benefitted from numerous suggestions from John T. Amenyo of ANS, Joel Halpern of Newbridge, and Andy Malis of Ascom-Timplex.

## Authors' Addresses:

Robert G. Cole
AT&T Bell Laboratories
101 Crawfords Corner Road, Rm. 3L-533
Holmdel, NJ 07733
Phone: (908) 949-1950
Fax: (908) 949-8887
Email: rgc@qsun.att.com

David H. Shur
AT&T Bell Laboratories
101 Crawfords Corner Road, Rm. 1F-338
Holmdel, NJ 07733
Phone: (908) 949-6719
Fax: (908) 949-5775
Email: d.shur@att.com

Curtis Villamizar
Advanced Network & Services

| Documents | Summary |
|---|---|
| RFC-1483 | How to identify/label multiple packet/frame-based protocols multiplexed over ATM AAL5. Applies to any model dealing with IP over ATM AAL5. |
| RFC-1577 | Model for transporting IP and ARP over ATM AAL5 in an IP subnet where all nodes share a common IP network prefix. Includes ARP server/Inv-ARP packet formats and procedures for SVC/PVC subnets. |
| RFC-1626 | Specifies default IP MTU size to be used with ATM AAL5. Requires use of PATH MTU discovery. Applies to any model dealing with IP over ATM AAL5 |
| RFC-1755 | Defines how implementations of IP over ATM should use ATM call control signaling procedures, and recommends values of mandatory and optional IEs focusing particularly on the Classical IP model. |
| IPMC | Defines how to support IP multicast in Classical IP model using either (or both) meshes of point-to-multipoint ATM VCs, or multicast server(s). |
| NHRP | Describes a protocol that can be used by hosts and routers to determine the NBMA next hop address of a destination in ''NBMA connectivity'' of the sending node. If the destination is not connected to the NBMA fabric, the IP and NBMA addresses of preferred egress points are returned. |

Table 5: Summary of WG Documents

100 Clearbrook Road
Elmsford, NY 10523
Email: curtis@ans.net

# References

[1] R. Atkinson. Default IP MTU for use over ATM AAL5. Request for Comments (Experimental) RFC 1626, Internet Engineering Task Force, May 1994.

[2] R. Braden and J. Postel. Requirements for Internet gateways. Request for Comments (Standard) RFC 1009, Internet Engineering Task Force, June 1987. Obsoletes RFC-985.

[3] R. Braden, J. Postel, and Y. Rekhter. Internet Architecture Extensions for Shared Media. Request for Comments (Informational) RFC 1620, Internet Engineering Task Force, May 1994.

[4] ATM Forum. *ATM User-Network Interface Specification.* Prentice Hall, September 1993.

[5] J. Garrett, J. Hagan, and J. Wong. Directed ARP. Request for Comments (Experimental) RFC 1433, Internet Engineering Task Force, March 1993.

[6] J. Heinanen. Multiprotocol Encapsulation over ATM Adaptation Layer 5. Request for Comments (Proposed Standard) RFC 1483, Internet Engineering Task Force, July 1993.

[7] J. Heinanen and R. Govindan. NBMA Address Resolution Protocol (NARP). Request for Comments (Experimental) RFC 1735, Internet Engineering Task Force, December 1994.

[8] M. Laubach. Classical IP and ARP over ATM. Request for Comments (Proposed Standard) RFC 1577, Internet Engineering Task Force, January 1994.

[9] J. Mogul and S. Deering. Path MTU discovery. Request for Comments (Draft Standard) RFC 1191, Internet Engineering Task Force, November 1990.

[10] M. Perez, F. Liaw, D. Grossman, A. Mankin, and A. Hoffman. ATM signalling support for IP over ATM. Request for Comments (Informational) RFC 1755, Internet Engineering Task Force, January 1995.

[11] International Telegraph and D. Mills. Exterior Gateway Protocol formal specification. Request for Comments (Historical) STD 18, RFC 904, Internet Engineering Task Force, April 1984.

# A Potential Interworking Scenarios to be Supported by ARP

The architectural model of the VC routing protocol, being defined by the Private Network-to-Network Interface (P-NNI) working group of the ATM Forum, categorizes ATM networks into two types:

- Those that participate in the VC routing protocols and use NSAP modeled addresses UNI 3.0 [4] (referred to as private networks, for short), and

- Those that do not participate in the VC routing protocol. Typically, but possibly not in all cases, public ATM networks that use native mode E.164 addresses UNI 3.0 [4] will fall into this later category.

The issue for ARP, then is to know what information must be returned to allow such connectivity. Consider the following scenarios:

- Private host to Private Host, no intervening public transit network(s): Clearly requires that ARP return only the NSAP modeled address format of the end host.
- Private host to Private host, through intervening public networks: In this case, the connection setup from host A to host B must transit the public network(s). This requires that at each ingress point to the public network that a routing decision be made as to which is the correct egress point from that public network to the next hop private ATM switch, and that the native E.164 address of that egress point be found (finding this is a VC routing problem, probably requiring configuration of the public network links and connectivity information). ARP should return, at least, the NSAP address of the endpoint in which case the mapping of the NSAP addresses to the E.164 address, as specified in [4], is the responsibility of ingress switch to the public network.
- Private Network Host to Public Network Host: To get connectivity between the public node and the private nodes requires the same kind of routing information discussed above - namely, the directly attached public network needs to know the (NSAP format) ATM address of the private station, and the native E.164 address of the egress point from the public network to that private network (or to that of an intervening transit private network etc.). There is some argument, that the ARP mechanism could return this egress point native E.164 address, but this may be considered inconsistent for ARP to return what to some is clearly routing information, and to others is required signaling information.

In the opposite direction, the private network node can use, and should only get, the E.164 address of the directly attached public node. What format should this information be carried in? This question is clearly answered, by Note 9 of Annex A of UNI 3.0 [4], vis:

> "A call originated on a Private UNI destined for an host which only has a native (non-NSAP) E.164 address (i.e. a system directly attached to a public network supporting the native E.164 format) will code the Called Party number information element in the (NSAP) E.164 private ATM Address Format, with the RD, AREA, and ESI fields set to zero. The Called Party Subaddress information element is not used."

Hence, in this case, ARP should return the E.164 address of the public ATM station in NSAP format. This is essentially implying an algorithmic resolution between the native E.164 and NSAP addresses of *directly* attached public stations.

- Public network host to Public network host, no intervening private network: In this case, clearly the Q.2931 requests would use native E.164 address formats.
- Public network host to Public network host, intervening private network: same as the case immediately above, since getting to and through the private network is a VC routing, not an addressing issue.

So several issues arise for ARP in supporting arbitrary connections between hosts on private and public network. One is how to distinguish between E.164 address and E.164 encoded NSAP modeled address. Another is what is the information to be supplied by ARP, e.g., in the public to private scenario should ARP return only the private NSAP modeled address or both an E.164 address, for a point of attachment between the public and private networks, along with the private NSAP modeled address.