Routing in IP over ATM Networks

Jussi Vähäpassi
ICL Data Oy
P.O.BOX 458, 00101 Helsinki, Finland
jussi.vahapassi@icl.fi

Abstract

This article gives an overview of several IP over ATM technologies. Their impact on IP routing protocols has been quite small. The Classical IP over ATM and LANE are the first overlay techniques developed. Their limitations were soon recognized and work began to create more sophisticated and scalable solutions. Some of the proposed methods never gained success. For quite a while the IP and ATM were treated as two totally separate protocol layers, and the nature of these protocols made the co-existence inefficient for both. Fast techniques emerged that enabled the creation of ATM paths based on IP traffic flows. Quite many vendors developed their version of fast IP switching of which each had its merit, but luckily quite soon a common platform known as MPLS was developed. Finally, the emergence of IPv6 seems to be delayed over and over again, but its development work goes on.

1 Introduction

Less than ten years ago ATM was seen as a technique that would extend to speeds that current technologies, such as Frame Relay or SMDS, would never reach. The first step was the 155 Mbit/s bit rate, which was enough for routers to handle. Quite soon 622 Mbit/s ATM interface was developed, but warnings were seen that scalability of ATM is a serious issue. And indeed, this seems to be the limit that cannot be easily broken even today.

Carriers and ISPs did not see the benefit of ATM in their backbones. The statistical multiplexing gain brought by ATM was not useful and cell tax wasted the scarce bandwidth of the backbones. The PPP over SONET [1]became the dominant technology in those networks where the ATM benefits were not available. ATM became more of an access and LAN technology where speeds are modest and statistical multiplexing is of use.

But in LANs legacy switches became faster and the deployment of gigabit Ethernet seems override ATM in local area networks. Currently 10 Gbit/s interface is under development while ATM does not scale. And ATM LANE technology was never quite scalable or robust, and the native ATM applications that would take use of ATM benefits never appeared since the upgrade step was too big. It looks like access networks technologies, such as residential broadband, still go with ATM.

Fast IP switching technologies gave boost to routing performance. Commonly these technologies by various vendors utilized ATM switching speed by separating IP routing and forwarding functions. The various implementations have come together under the umbrella of MPLS. Recently the IP routing and forwarding speeds have increased remarkably, and the importance of MPLS has decreased. However, MPLS has benefits that legacy IP routers do not implement efficiently today: traffic engineering, CoS support and VPN support [2]. It remains to be seen whether the all-mighty IP routers can implement these efficiently which would cause even MPLS to vanish.

2 Traditional IP over ATM

The traditional methods of transporting IP over ATM networks separate ATM and IP layers from each other quite strictly. This suits the TCP/IP layering, where the network protocol works on top of link layer protocols.

2.1 LAN Emulation (LANE)

LANE [3] is in essence a protocol for bridging across ATM. It is solely a Local Area Network protocol, like Ethernet, and it enabled the fast deployment of ATM to local areas by supporting a variety of network layer protocols, since native ATM applications were not deployed as fast as some had hoped (as it later turned out, applications didn't emerge at all).

Since it is a LAN technology routers must be used to connect different subnets together. LANE itself does not care what the network protocol is; it transports all protocols equally just like other LAN technologies. LANE requires no modifications to upper layer protocols to make them work in ATM networks. Because ATM supports multicast and broadcast weakly, this functionality is implemented with a dedicated server.

It was originally envisaged that LANE will appear in workstation and server network interface cards, and ATM attached LAN switches and routers. While NICs were only moderately deployed, LANE had some success in switches and routers.

Although ATM switches are often used to house some components of LANE functionality, generally ATM switches need not be aware of the LANE. This is because it builds upon the overlay model in which LANE operates transparently through ATM switches. Only the main signaling components are utilized.

The basic function of LANE is to carry out the MAC address to ATM address resolving. Once the resolving is done, it is the native ATM layer job to create the path between end systems. Additional complexity comes from the need to create separate LAN's into one ATM network, called emulated LAN's (ELAN). While LANE can support Ethernet and Token Ring ELANs, it is not possible to directly forward frames between these ELANs, but instead a router must be used.

The scalability of LANE is limited by the size of a broadcast domain. Just as in Ethernet networks, if the network size becomes too big, the broadcast traffic will become a problem. After all, LANE is just another local area network technology.

Being just another local area network that tries to hide the properties of ATM from higher layer protocols, LANE does not support ATM QoS at all. End stations can not utilize ATM service classes.

When LEC (LANE Client) wants to make connection to another LEC after having resolved its ATM address, usually a SVC is created. The creation of SVC relies on ATM routing protocol if the network consists of several ATM switches. The routing is totally out of scope for LANE, and it relies on active ATM support on this. Typically static routes are used (IISP [13]), or a dynamic routing protocol (PNNI [12]) may distribute routing information. Most often LANE network consists of one ATM switch, so the routing problem becomes simple.

The biggest advantage of LANE is that it can support ATM and legacy networks in the same LAN. It however generates additional traffic since in addition to IP-to-MAC ARP a similar resolution must be done between MAC and ATM addresses.

2.2 Classical IP over ATM

The classical model is known by the RFC1577, but it is now obsoleted by RFC2225 [4]. The IP subnetting is preserved so that communication between hosts in different subnets goes via a router. Hence the name "Classical Model": even if the end hosts are in the same ATM cloud and direct virtual connection could be established, there may be extra hops in the path if the hosts are in different subnets. The IP subnets are called Logical IP Subnets (LIS).

The packet encapsulation has been defined in RFC1483 ([9], now obsoleted by [10]), which also takes into account the saving of virtual connections by multiplexing different protocols into a single VC. The LLC/SNAP encapsulation defined in RFC1483 is the most common encapsulation in IP over ATM networks.

The basic functionality of the classical model is to resolve ATM to IP address mapping. Each LIS contains an ATMARP server, and all clients are configured with the address of this server. Once a client joins the LIS it registers itself with ATMARP server and thus the server becomes aware of the client. When other clients want to communicate with it they ask ATMARP server for the client's ATM address and can thus establish a direct data VC to it. ARP entries age out in clients in 15 minutes and in server in 20 minutes.

RFC2225 brings two major improvements to the model. Inverse ARP need not be used in SVC environment but instead server may generate its ARP table by examining messages from client. Also the clients may be configured with several ATMARP server addresses, from which to choose. While this does not produce load sharing between servers and other intelligence, it brings resilience to the service since clients have the possibility to use another server if one is down.

From routing point of view the situation is same as with LANE. The network can operate in PVC or SVC environment. When using SVCs Classical IP over ATM relies on lower layer ATM protocols to route the SVCs across the network. The IP layer routing protocols are usable, and since this is an overlay model the IP and ATM layer protocols are independent from each other.

The advantage of Classical IP over ATM is that it simplifies the protocol stack when the LAN contains only ATM switches, and it provides efficient transport for IP traffic.

The disadvantage is the need for extra hop between different IP subnets even if they are in the same ATM network.

2.3 Next Hop Resolution Protocol (NHRP)

The Classical Model's extra hop problem is being solved in NHRP by the IETF IP Over Non-Broadcast Multi-Access (NBMA) Network (ION) working group. NHRP supports cut-through routing, and uses routers mainly to for end hosts address resolving. The protocol is not technology specific, but can operate over any NBMA network, such as Frame Relay or ATM [5].

If nodes are in the same NBMA network, they can establish direct connection between each other, as opposed to Classical IP over ATM where direct connectivity is possible only if the nodes belong to the same IP subnet. NHRP can even resolve addresses of hosts, which are not directly connected to the NBMA network. In this case the address of the egress router is used. The connectivity can be limited administratively between NBMA Logical subnetworks, so that some form of "policy firewalling" is supported. Within each logical NHRP subnetwork there are Next Hop Servers (NHS) that provide NHRP service to a set of hosts.

The main task of routers in NHRP is not to forward data packets. Instead they serve NHRP Resolution Requests. If a client does not know the destination address, it sends the Resolution Request to its NHS, which usually is a router. The NHS will have some knowledge of the network based on higher layer routing protocols and will either answer the request directly or forward the request to another NHS. The Request Reply is returned the same path backwards so that all intermediate NHSs learn the new address and can cache it. This is especially useful if the reply contains an address that serves a whole IP subnet.

When the client receives the reply it may create a direct connection to destination. Since the shortcut is not necessarily created to the end host but instead to an egress router, the path may not be optimum end-to-end. This is because the higher level routing protocol does not have full information on the NBMA network topology. This fact also leads to difficult issues in QoS routing and in some circumstances creates a potential stable routing loop.

While the request is being sent to another NHS there is a choice to be made whether to forward the actual data packet also or to buffer it to wait for the resolution reply. The forwarding across routers is the quickest solution and introduces smallest latency, but it may lead to packet disordering when a shortcut becomes available.

Since two routers in NHRP network can create a direct connection between each other across domain boundaries that is used only for data forwarding and not routing protocol updates, there is a possibility of stable routing loops to be formed. This may happen between multi-homed networks. Traditionally this is avoided by requiring that routing updates should be sent across all paths that data also flows. This is a risk in administrative domain boundaries where routing protocol metrics is lost. Options have been discussed widely to overcome this problem, but a final solution is not found. The safest option is to limit to a classical model, i.e. using routers to forward data packets across domain boundaries. NHRP is not a replacement to routing protocols, nor does it supplement them. The information obtained via NHRP should not be used for routing. Instead, NHRP is an address resolution protocol, and may be used to supplement RFC2225 ARP. [6]

NHRP does solve the extra hop problem. But extra hops cannot be eliminated between administrative domain boundaries, so the creation is ubiquitous ATM infrastructure cannot be done by NHRP.

Currently NHRP focuses on unicast routing. There is serious doubt whether NHRP can ever support scalable multicast routing. NHRP does not take into account layer 2 switches or virtual LAN's, but instead assumes that all devices are either routers or hosts.

2.4 Multi-Protocol Over ATM (MPOA)

MPOA is a joint effort of ATM Forum and IETF to accomplish internetworking of ATM networks with legacy subnetworks. The Multi-protocol over ATM Specification, Version 1.1 has been approved in May, 1999 [7]. The building blocks are:

- LANE
- NHRP
- MARS for multicast connectivity
- UNI signaling
- RFC1483 encapsulation
- Spanning tree for VLAN support

As seen above, the components of MPOA are defined separately. The only thing new in MPOA is that the framework brings together existing ATM and legacy internetworking elements. MPOA supports detection of flows and creation of shortcuts based on this information. NHRP has been extended to include tags, which help edge devices to detect flows.

While NHRP is designed to be non-ATM specific, MPOA will make some modification to make it more ATM specific and less IP specific. MPOA also takes into account layer 2 connectivity in that it supports virtual LANs and ATM capable switches. This is accomplished by LANE.

MPOA supports Layer 3 switching equipment, because the packet routing and packet forwarding functions have been separated into different functional groups. Route servers can handle the software-intensive routing protocol updates and route processing, while packet forwarding can be done in inexpensive but efficient layer 3 switches. A protocol is being developed to distribute the routes to forwarding devices (the "MPOA protocol").

MPOA clients are capable of setting up a direct data connection across the ATM network based on layer 3 addresses. To avoid problems similar to NHRP routing loops, this address must be that of the "final node" in the ATM network and not the address of an intermediate node in the ATM network.

The most important benefit of MPOA is that it supports VLANs. The difficulty of MPOA lies in implementation details since it tries to merge together a number of technologies and so it must make some modifications to each of them. MPOA suffers from the same limitations in network topology as NHRP in that multihomed networks may suffer from persistent routing loops.

3 IP switching

The IP switching is a common term to describe the approach of reducing the longest-match destination prefix lookup performance bottleneck of traditional routers. This term is used generally despite the fact that one of the pioneers in this area, Ipsilon, used this term for its own specific technology.

The main idea is to speed up the table lookup by assigning a "tag" or "label" to packets. A fixed length tag is faster to process than IP prefixes, and "tag swapping" within a switch is easily done in hardware.

As a new generation of routers has emerged that are capable of wire-speed routing in gigabit speeds, the advantage of fast label swapping is no longer there. Instead, now the most important benefit in IP switching technologies is the possibility for traffic engineering.

There are two main streams in IP switching: flow based switching and route based switching. The first approach uses some criteria in routers to determine whether a packet stream is long lived. If it is then a label is assigned and fast switching mode is entered for the rest of the flow. The attributes used in determining a flow are source IP address, destination IP address, protocol and port numbers. Flow based switching is generally not considered a scalable solution since the number of flows is quite high in large networks. The latter assigns tags based of IP routing information, not on individual flows, and is considered to be a more scalable solution.

This way paths are created through the network. The paths can usually also be created through management action so that traffic engineering can be accomplished.

3.1 Cisco Tag Switching

Cisco's tag switching does not rely solely on ATM layer, but it should be applicable to most lower layer technologies. [8]

Tag switching is a flow-based technology. A tag is assigned to a packet as soon as first packet of a flow enters a tag switching capable equipment. When a tag switch receives a packet with a tag, the switch uses the tag as an index in its Tag Information Base (TIB). Each entry in the TIB consists of an incoming tag, and one or more sub-entries of the form (outgoing tag, outgoing interface, outgoing link level information). If the switch finds an entry with the incoming tag equal to the tag carried in the packet, then for each (outgoing tag, outgoing interface, outgoing link level information) in the entry the switch replaces the tag in the packet with the outgoing tag, replaces the link level information (e.g. MAC address) in the packet with the outgoing link level information, and forwards the packet over the outgoing interface.

The tags are quite similar to ATM VPI/VCI field, but they carry more information. This makes it a bit difficult to map tags directly to ATM VPI/VCI in case of ATM transport. In other technologies where the packet size is not as much restricted the mapping is more straightforward. Also tags may be stacked which allows information hiding, such as tunneling, VPNs and overlapping private address spaces.

Routing information is stored in Forwarding Information Base (FIB). Devices exchange this information by using Tag Distribution Protocol (TDP).

Since tag switching is destination based, the underlying technology must be capable of flow merging. Multiple flows, which are going to same egress point and share QoS parameters must be merged into the same tag switched path. ATM switches are not capable of VC-merging. This is because there is no multiplexing information within cells. ATM equipment expect that all cells with same VPI/VCI information belong to the same stream. The requirement to map a single Tag Switched stream into multiple ATM virtual circuits is hardly scalable to Internet levels.

TTL decrementing is a method of reducing problems caused by routing loops. However, ATM switches are not capable of decrementing the TTL field since they assume that ATM routing protocols prevents the formation of loops totally. But this is not the case with IP routing protocols; they accept transient routing loops, so decrementing the TTL field must be carried out in advance in edge router before entering ATM switched cloud.

3.2 IBM ARIS

Aggregate Route-Based IP Switching (ARIS) is IBM's effort to introduce fast switching into IP networks. It is route based switching technology, like MPLS but unlike tag switching. Packets from any ingress point forming a leaf on the route tree, and intended for that egress point's destination prefix, are switched and merged to the root egress ISR.

ARIS is remarkable in one respect: ARIS ISRs (Integrated Switch-Router) support VC merging by encouraging the use of ATM switches that have been specifically designed to buffer AAL5 PDUs. Packets arriving with different VP/VCs can be forwarded onto a single VP/VC (merged) by being retransmitted sequentially once an entire datagram has been received, without any cell interleaving.

At startup an ARIS network establishes switched paths to all egress points. This happens regardless of any traffic. Each egress ISR starts a path establishment sequence. ARIS network is guaranteed to be loop-free, since the establish-message appends an ISR ID, and thus ignore the message if it contains its own ID (much like BGP-4 AS_PATH attribute).

ARIS supports source routing and multicasting. Switched path information is soft state, and KeepAlive messages are used to maintain state if no traffic is present.

3.3 Ipsilon IP Switching

IP switching is dedicated to IP and ATM. The two components are IP Switch Controller and ATM switch. ATM switch is used purely to switch ATM cells, and the IP Switch Controller controls the labels.

IP Switch Controller handles both routing and forwarding. As soon as the controller notices that there is a flow, it requests that the upstream switch dedicate a virtual circuit for it. When the upstream switch does this the ATM switch is instructed to switch it to the IP forwarding engine via this new VC rather than the common, connectionless VC used for incoming default IP traffic. Now when the downstream node has also identified this as a persistent flow, and requested (via IFMP) that a unique VC be set up to carry it, the IP Controller recognizes the opportunity for cut-through. This flow enters the ATM switch on a unique VC and is switched up to the IP Controller on this VC; forwarded normally there via another unique outgoing VC back to the ATM switch, and then out to the downstream node. The IP Controller instructs the switch to logically connect the incoming VC with the outgoing VC within the switch fabric, without being routed up to the IP Controller, and the cut-through route is established. At this point, the IP Controller no longer sees this flow at all. Flows are soft state, and they are deleted when there is no more traffic.

Some additional performance enhancement is brought by encapsulation. As long as flows go via Switch Controller the encapsulation is RFC1483 LLC/SNAP [9][10]. But as cut-through path is established, the encapsulation is changed to RFC1483 VC-multiplexed. More important than performance is that security is enhanced remarkably.

Traditional IP routing protocols are usable without modifications, but they can not take use of ATM's QoS offerings. Scalability of this solution has been unclear. It requires a VC for each flow, thus creating a possibility of VC starvation.

3.4 CSR

The Cell Switch Router (CSR) architecture has been developed in Japan and has been brought into commercial use by Toshiba [11].

The purpose of CRS is to connect ATM LANE and Classical IP over ATM networks. Connection between ATM LIS and non-ATM networks goes via standard routers. In functionality the CSR is very similar to Ipsilon IP Switching in that controller decides when a flow is detected and establishes a direct connection into ATM switch. CSR supports also non-IP protocols, and unlike Ipsilon IP switching, it can switch based in IP subnet information.

CSR supports QoS in that RSVP packets can trigger ATM VC setup, but ATM QoS is available only between CSRs.

4 PNNI

Private-NNI (P-NNI) is not an IP over ATM technique in itself, but purely a complex, highly scalable ATM routing and signaling protocol. It supports QoS routing, hierarchical networks, source routing and scalability. The P-NNI V1.0 specification was approved in March, 1996 [12].

The predecessor of P-NNI is Interim Inter-Switch Signaling Protocol (IISP), also known as P-NNI Phase 0 which was developed because it was seen that P-NNI specification would take a long time to finish. IISP supports only static routes and is not interoperable. [13]

P-NNI is based on link-state algorithm. Topology information (including information about nodes, links, addresses) is flooded through the network. Network resources are defined by metrics and attributes (delay, available bandwidth, jitter, etc.), which are grouped by supported traffic class. Since some of the metrics used will change frequently (e.g., available bandwidth), threshold algorithms are used to determine if the change in a metric or attribute is significant enough to require propagation of updated information.

It has been proposed that for IP to be able to better utilize the ATM features and network knowledge, integrated IP and ATM routing should be established. Two attempts address this: P-NNI Augmented Routing (PAR) and Integrated PNNI (I-PNNI).

4.1 PAR

IP routing protocols generally have little information about the ATM network topology and QoS capabilities. P-NNI Augmented Routing (PAR) makes the ATM edge routers run P-NNI so they will become aware of the ATM network topology. ATM switches do not need to know about IP; they run P-NNI as usual. The PNNI Augmented Routing (PAR) Version 1.0 specification was approved by ATM Forum in January 1999 and PAR is still under development [14][15].

ATM switches see edge routers as ATM capable devices and are capable of setting up SVCs to them as necessary. However, edge routers show as "restricted transit nodes" to switches, so transit SVCs can not be created through routers.

P-NNI protocol allows the carrying of information elements that need not be interpreted by ATM switches. Instead, edge routers can carry routing information in these TLV-encoded fields through ATM cloud.

Proxy PAR (PPAR) is a protocol that allows ATM attached devices to interact with PAR-capable switches without executing PAR themselves. This allows easy implementation in e.g. IP routers, since Proxy PAR client is much lighter than full functioning PAR client [16].

4.2 I-PNNI

Integrated P-NNI (I-PNNI) is more ambitious than PAR, since it tries to solve the integrated routing problem by making all ATM switches and ATM edge routers to run the same protocol which is aware of both the IP network and ATM topology [17][18].

The major architectural factor in I-PNNI is that every router and switch must be I-PNNI-aware. This way the power and scalability of P-NNI can be brought to IP world. Routers running I-PNNI would support a hierarchy similar to ATM switching systems.

I-PNNI certainly held great promise as a routing protocol for the Internet with QoS. But the problem is that the current core is a routed core, and introducing switching and changing the routing protocol is a heavy task which is not done easily. Although some vendors announced support for I-PNNI, the ATM Forum has abandoned it.

5 MPLS

MultiProtocol Label Switching (MPLS) is one of the new networking techniques which aims at combining the flexibility of the IP protocol routing with cell switching technology. It is one of the hottest topics in the industry at the moment. MPLS is a "merger" of most of the older IP switching technologies in that it seems to contain many features of the former IP switching technologies [19] [20].

MPLS consists of three components:

- IP routing protocol, e.g. OSPF, BGP-4, is used at the edge and in the core to find routes for paths through the MPLS network
- IP forwarding based on traditional longest prefix match is used at the edge of the network
- Label based forwarding is used in the core, which uses hardware based exact match for rapid forwarding.

MPLS is not dedicated to any layer 2 technology. It uses labels that are native to the media, e.g. VPI/VCI fields in ATM and DLCI field in Frame Relay. All streams that are assigned the same forwarding equivalence class (FEC) share the same destination, path and CoS/QoS features. The path is called Label Switched Path (LSP), and a switch/router using MPLS is called Label Switching Router (LSR). Label Distribution Protocol (LDP) is used to share protocols between LSRs.

Packets are assigned a label at the entry to a MPLS domain, which is often the core backbone of a provider, and are switched inside the domain by a simple label lookup. The labels determine the quality of service the packet gets. The labels are stripped off at the egress router and packets are routed in conventional fashion to the final destination.

Labels can be stacked, where a packet carries more than one label. This allows tunneling of packets. Labels are popped in "Last in first out" method. A router always makes the forwarding decision based on the topmost label. The labels can then be popped, swapped or added. The topmost label(s) can be the native address field, e.g. in ATM the VPI/VCI field. But the next labels in the stack must be encapsulated into a "shim header", an extra layer header in AAL5 PDU.

Traffic engineering, one of the main MPLS features, allows one or more streams to be forwarded according to a pre-defined path[21]. It gives the opportunity to tailor and balance traffic in the network so that standard routing information can be overridden and well-defined streams can be routed differently. This is accomplished by source routing.

5.1 Route selection

MPLS supports both hop-by-hop routing and explicit routing (source routing). With explicit routing flexibility is gained, where the routing can be based on QoS requirements or other policies. The route selected need not be shortest path.

In hop-by-hop routing the path is chosen at each Label Switching Router at packet arrival as in legacy IP networks. The routing process is distributed so that all hops participate in the routing decision. Traffic engineering is very difficult.

With explicit routing the ingress router selects the LSP explicitly that the packet takes. The LSP can be specified partly (loosely explicitly routed) or it can be declared completely (strictly explicitly routed). The path specified this way is called Explicitly Routed LSP (ER-LSP). The routing burden is concentrated to the ingress nodes. Sophisticated mechanisms are required to communicate the constraints to the routing process. This is done manually or via some automation.

5.2 Label Distribution Protocol

The purpose of LDP is to communicate labels and label-to-FEC binding between adjacent routers. Labels do not have significance over the whole network, but instead they are significant only locally. Depending on the network the LDP can either be included as part of existing routing protocol or it can be developed as a dedicated protocol.

FECs can be created in advance (independent LSP control). The downstream LSR creates a FEC-to-label binding and communicates this to upstream LSR, which inserts this into its forwarding tables. This creates labels with less delay but may require separate loop detection method. If the LSR decides to create a label for every destination can reach, a depletion of labels can occur on systems where label space is limited.

Also bindings can be generated on-demand (ordered LSP control) where LSR does not allocate label to a destination until requested to do so, and it will not reply to that request until it has received a label from its downstream peer. This creates some delay before packets can be forwarded along the LSP but does not create loops. This method is used for explicit routing and multicast.

Both methods are supported in the same network at the same time. There may be more than one possible LSP. What happens to additional LSPs depends on the mode: In liberal retention mode the extra labels are kept for possible future use in case the primary LSP becomes unavailable. Then another path can be taken into use quickly. This consumes resources since a lot of unused labels must be kept in memory. In conservative retention mode LSR only maintains bindings received from valid next hop, and the extra labels are released the moment they are received.

However, LDP can only follow the IP forwarding tables and thus can not support traffic engineering. To address this shortcoming an extension to LDP, called CR-LDP, has been introduced to better support traffic engineering. Also RSVP with extensions tackle the requirement for traffic engineering. It remains to be seen to which extent plain LDP is needed. A good comparison of RSVP and CR-LDP is presented in [22].

5.3 CR-LDP

With constraint-based LSP (CR-LDP) the path is formed along an explicitly defined route (source route). When a LSR receives a CR-LDP request message, it does not follow the IP forwarding table as LDP would do, but instead it follows the route as instructed in the message. This forms a constraint-based LSP (CR-LSP)

The ability of CR-LDP to compute explicit routes dynamically is not yet defined, but it has been proposed that OSPF or IS-IS link state advertisements (LSA) contain dynamic bandwidth reservation information. With this mechanism CR-LDP can reserve bandwidth which is checked against available bandwidth. The reservation is made and the new available bandwidth is flooded to other nodes via OSPF or IS-IS extensions.

The constraints that can be used are few, usually bandwidth and number of hops, or bandwidth and delay. Other QoS elements such as jitter and packet loss are not readily supported.

The traffic parameters that can be used to describe the path behavior are:

- Peak Data Rate (PDR)
- Peak Burst Size (PBS)
- Committed Data Rate (CDR)
- Committed Burst Size (CBS)
- Excess Burst Size (EBS)

PDR and PBS define the maximum rate at which data should be sent to CR-LSP. The edge function should police that this is not exceeded. CDR and CBS are the rates to which MPLS domain commits and resources are reserved. EBS is an additional limit to burst sized that should be regulated at the edge. The specification of traffic parameters is optional. If the fields are left empty then best effort service is used.

DS-field can be used to define the packet behavior, e.g. drop precedence. The packet marking is done at the edge node so services can be created.

CR-LDP pre-emption (bumping) can be used to steal resources from lower priority LSPs. This may be used during times of failure so that more important LSPs obtain resources before less important LSPs.

5.4 RVSP for Traffic Engineering

CR-LDP is a new protocol, which borrows much of its code from LDP [23]. Another choice for use in traffic engineering is RSVP with extensions. Some favor this since a lot of work has been done earlier with RSVP, while CR-LDP is a totally new protocol. [24]

Unlike CR-LDP, which relies on TCP for reliable transport, RSVP uses IP datagrams directly, and thus must implement reliability itself.

Quite a few vendors are going to support RSVP extensions for traffic engineering. It is beneficial if the platform lacks support for TCP, but this is not a major obstacle. RSVP requires slight modifications to IP processing, since all RSVP packets must be directed to RSVP process regardless of their destination IP address. This requirement also makes it impossible to use IPSec.

There are certain doubts whether the RSVP is scalable for networks with high number of LSPs. The reservation is soft state and keep-alive messages flow with configurable interval between adjacent nodes. This load may be reduced by bundling many LSP refresh messages into one [25]. With this in use the RSVP, as well as CR-LDP, scales with the number of LSRs, instead of the number of LSPs.

5.5 Why MPLS?

MPLS fits into IP routing paradigm in a way that the core functionality need not be changed. The problem with many other technologies is that moving from existing routed core to e.g. ATM switched core is too big a step.

Fast forwarding used to be one major selling point for MPLS. This is no longer a major issue since new generation of routers have appeared that are capable of wire speed routing similar to fastest switching platforms.

The most important reason for MPLS deployment is traffic engineering. The routing can be based on constraints (CR-LDP). Sometimes certain packets are desired to be routed along certain specific paths which are decided beforehand or when the packet enters the network. In IP, this requires source routing. But in MPLS this is easily done with the help of source routing that bypasses routing protocols.

VPNs are supported by tunneling. Thus security is sufficient. Also overlapping private IP network spaces are sometimes a problem which can be helped with tunneling mechanism.

MPLS supports partly QoS features. It can make use of ATM QoS classes, and IP routing protocols that take QoS into use can be used, such as QOSPF.

6 IPv6 over ATM routing

IPv6 has been primarily designed to solve the address shortage problem of IPv4 networks. Several other techniques have extended the estimated lifetime of IPv4 and thus the urgency of IPv6 has reduced. The transport of IPv6 over ATM networks does not bring anything new to ATM networks in routing sense. The same modern routing protocols are still as usable as they have always been.

The basic problem in adapting IPv6 to ATM networks is that IPv6 relies very heavily on connectionless multicast capabilities. The Neighbor Discovery protocol, which IPv6 uses to perform neighbor and router discovery, assumes that if the data link address of a certain node is not available, it still can be reached by sending a multicast message. However, ATM networks do not provide connectionless multicast link services at the datalink. Thus, some mechanism to provide a multicast service for use by IPv6 must be provided if all the IPv6 discovery protocols are to be preserved on ATM networks. In PVC environment multicast services are not needed since multicast and broadcast operations collapse down to an ATM level unicast operation. In SVC environment IPv6 relies on MARS servers to provide the multicast functionality [26].

Additionally, IPv6 differs from IPv4 in that address resolution and address configuration are located in the network layer rather than the datalink layer. That is, the ND protocols are an integral part of the IPv6 network layer (ICMPv6), and any mechanism that is used to adapt ATM to IPv6 must deal with the Neighbor Discovery protocols. This is in contrast to IPv4 where the address resolution protocols are not part of the base IP protocols but are part of each individual datalink layer (i.e., ARP for broadcast media, ATMARP or NHRP for ATM). In IPv4 new datalink layers could define their own address resolution protocols as necessary (as was done with ATMARP) since this function is left to the datalink. Thus, new datalinks could be added without affecting the IPv4 network layer. In IPv6 all datalinks must handle IPv6 Neighbor Discovery packets and use them for address resolution, router discovery and address configuration. Not using Neighbor Discovery would require modifying the IPv6 network layer to accommodate a specific datalink. For PVC environment the dynamic shortcuts are not supported, thus NHRP services are unnecessary. In SVC environment NHRP is used for seeking out the NBMA identities of IP interfaces that are logically distant in an IP topological sense [27].

Further, IPv6 provides for some extra features not in IPv4. One of these features is address autoconfiguration. Address autoconfiguration allows a host to configure one or more addresses per interface automatically and without explicit system administration. Another IPv6 feature is security. IPv6 has been defined with network layer security features as part of the base protocol. These security features are applied to the discovery and configuration protocols since these protocols are defined at the network layer. All current IPv6 security mechanisms will work without modification for ATM. This includes both authentication and encryption for both Neighbor Discovery protocols as well as the exchange of IPv6 data packets. Finally, IPv6 includes an address architecture, which provides for address scoping for both unicast and multicast addresses. This addressing architecture is also maintained for IPv6 over ATM.

7 Conclusion

The marriage of IP and ATM has never been easy. The basic difference between connectionless (IP) and connection-oriented (ATM) protocol is not easy to accommodate. Quick solutions appeared that only concentrated on IP packet forwarding in limited environment, and did not provide additional services at all. Faster techniques emerges, and slowly it looks like the CoS features will exist in IP/ATM networks.

Through all development IP routing protocols have not needed major changes. To them the ATM network looks like another NBMA network with its point-to-point links. A serious attempt to converge the IP and ATM routing is Integrated PNNI. But the community was not ready for such "forklift upgrade". The work with I-PNNI has stopped, but it may continue when the time is right.

Currently the work focuses on MPLS in IETF. The familiar protocols such as RSVP and OSPF are taken into use and extended to suit MPLS and CoS requirements. Their basic functionality and algorithms stay the same, however.

References

[1] A. Malis, W. Simpson: PPP over SONET/SDH, RFC2615, June 1999.

[2] C. Semeria: Multiprotocol Label Switching – Enhancing Routing in the New Public Network, White paper, Juniper Networks, September 1999, www.juniper.net.

[3] LAN Emulation over ATM 1.0, ATM Forum, January 1995, ftp://ftp.atmforum.com/pub/approved-specs/af-lane-0021.000.pdf

[4] M. Laubach, J. Halpern: Classical IP and ARP over ATM, RFC2225, April 1998.

[5] J. Luciani, D. Katz, D. Piscitello, B. Cole, N. Doraswamy: NBMA Next Hop Resolution Protocol (NHRP), RFC2332, April 1998.

[6] R. Cole, D. Shur, C. Villamizar: IP over ATM: A Framework Document, RFC1932, April 1996.

[7] The Multi-protocol Over ATM Specification, Version 1.1, ATM Forum, May 1999, ftp://ftp.atmforum.com/pub/approved-specs/af-mpoa-0114.000.pdf

[8] Y. Rekhter, B. Davie, D. Katz, E. Rosen, G. Swallow: Cisco Systems' Tag Switching Architecture Overview, RFC2105, February 1997.

[9] J. Heinänen: Multiprotocol Encapsulation over ATM Adaptation Layer 5, RFC1483, July 1993.

[10] Grossman, J. Heinanen: Multiprotocol Encapsulation over ATM Adaptation Layer 5, RFC2684, September 1999.

[11] Y. Katsube, K. Nagami, H. Esaki: Toshiba's Router Architecture Extensions for ATM: Overview, RFC2098, February 1997.

[12] P-NNI V1.0, ATM Forum, March 1996, ftp://ftp.atmforum.com/pub/approved-specs/af-pnni-0055.000.pdf

[13] Interim Inter-Switch Signaling Protocol, ATM Forum, December 1994, ftp://ftp.atmforum.com/pub/approved-specs/af-pnni-0026.000.pdf

[14] R. Callon, J. Jeffords, J. Drake, H. Sandick, J. Halpern: An Overview of PNNI Augmented Routing, ATM Forum / 96-0354, April 1996.

[15] PNNI Augmented Routing (PAR) Version 1.0, ATM Forum, January 1999, ftp://ftp.atmforum.com/pub/approved-specs/af-ra-0104.000.pdf

[16] P. Droz, T. Przygienda: Proxy-PAR, Work in progress, IETF ION working group, August 2000, www.ietf.org/internet-drafts/draft-ietf-ion-proxypar-arch-02.txt

[17] R. Callon, J. Jeffords, H. Sandick, J. Halpern: Issues and Approaches for Integrated PNNI, ATM Forum / 96-0355, April 1996.

[18] J. Jeffords: Integrated PNNI (I-PNNI) v1.0 Specification (working document), ATM Forum/BTD-PNNI-IPNNI-01.01, April 1997.

[19] IETF MPLS Working Group: www.ietf.org/html.charters/mpls-charter.html

[20] E. Rosen, A. Viswanathan, R. Callon: Multiprotocol Label Switching Architecture, Work in progress, IETF, August 1999, search.ietf.org/internet-drafts/draft-ietf-mpls-arch-06.txt

[21] C. Semeria: Traffic Engineering for the New Public Network, White paper, Juniper Networks, September 1999, www.juniper.net.

[22] P. Brittain, A. Farrel: MPLS Traffic Engineering: a Choice of Signaling Protocols, White paper, Data Connection Ltd, January 2000, www.datcon.co.uk

[23] IP Traffic Engineering for Carrier Networks: Using Constraint-Based Routing to Deliver New Services, White Paper, Nortel Networks, 1999

[24] C. Semeria: RSVP Signaling Extensions for MPLS Traffic Engineering, White paper, Juniper Networks, 1999, www.juniper.net.

[25] L. Berger, D. Gan, P. Pan, F. Tommasi: RSVP Refresh Overhead Reduction Extensions, Work in progress, IETF, March 2000, search.ietf.org/internet-drafts/draft-ietf-rsvp-refresh-reduct-03.txt

[26] G. Armitage, P. Schulter, M. Jork: IPv6 over ATM Networks, RFC 2492, January 1999.

[27] G. Armitage, P. Schulter, M. Jork, G. Harter: IPv6 over Non-Broadcast Multiple Access (NBMA) networks, RFC2491, January 1999.