ATM switching and IP Routing Integration: the Next Stage in Internet Evolution?

Paul Patrick White
Department of Computer Science
University College London
Gower Street
London
WC1E 6BT

email: p.white@cs.ucl.ac.uk
tel: +44 171 419 3701

(This paper appeared in the April 1998 Issue of IEEE Communications magazine)

Abstract

User demand for more bandwidth and QoS support has fuelled interest in the use of ATM as an underlying link-layer technology in the Internet. The challenge is how best to exploit the potential benefits of ATM while maintaining the inherent strengths of the IP layer that have made the Internet so successful. The suitability of various schemes with regard to meeting these goals is described. In particular we focus on recent work that tightly integrates the IP and ATM layers to produce hybrid(or integrated) ATM switch/ IP routers with very favourable characteristics.

Introduction

The Internet was conceived upon the principle of connectionless IP(Internet Protocol) datagram delivery whereby no per-session state needs to be setup in the network prior to sending a user's datagrams. Instead each IP datagram is routed hop by hop towards its destination according to the destination's globally unique IP address which is contained in the IP header. Each router that is traversed will examine this IP destination address and look for a match in its routing table in order to determine the correct outgoing interface for the next hop of the journey towards the destination. This connectionless model makes no assumptions about the underlying networks. Furthermore it does not implement call admission control or per flow resource reservation and consequently is unable to offer any QoS guarantees. The delivery service is termed 'best-effort' which means that each IP data flow is subject to an indeterminate level of packet loss, re-ordering, and delay, all of which increase with network load. On top of this core service, end-to-end reliability can be achieved through appropriate transport layer protocols such as TCP(Transmission Control Protocol) which uses such techniques as positive acknowledgement and retransmissions. The pooling of resources inherent in the traditional Internet philosophy is a key strength that ensures high utilisation of resources while overload results in graceful degradation of service rather than total collapse. In addition a connectionless model is very robust and handles an extremely wide range of failure scenarios without imposing a heavy signalling burden[1].

This classical model of IP has proven incredibly successful and the number of hosts on the Internet continues to double approximately every year. Within this overall growth rate there is also an increasing demand for multimedia applications that ask a lot more from the network than the more traditional types of Internet traffic such as File Transfer Protocol(FTP). In fact many of these multimedia applications have quite stringent Quality of Service(QoS) delivery needs in terms of packet delay, loss rate and minimum bandwidth. Furthermore as the World-Wide Web is increasingly used for business there is a growing number of users for whom delay-bounded access of information is especially important.

It is clear that if the Internet is to keep pace with such demands it must offer QoS support as well as increasing the bandwidth available to end-users. QoS support for IP flows at the IP layer is feasible by reserving resources on a per-flow basis which can be initiated by the user on demand using a reservation protocol such as RSVP[25]. Also, the advent of fibre-optic cables has resulted in a transmission medium with massive potential bandwidth capacity. However, the bottleneck arises at the communications nodes used to interconnect such media. The concept of IP routing was geared more towards flexibility rather than speed. Consequently, for a given cost, switches based on Asynchronous Transfer Mode(ATM) which was designed from the outset with high switching speeds in mind are able to operate at higher bit-rates than conventional IP routers. ATM achieves high bit rates through hardware switching of fixed-length cells with a small fixed-length header based on a one-to-one match with switching table entries. By contrast, forwarding decisions in an IP router were traditionally carried out in software based on a longest-match of the address prefix with entries in a routing table. In addition to ATM's ability to switch at high speed, current implementations also support QoS on demand so it may appear as though ATM solves all of the problems that the Internet is currently experiencing. In reality such an assumption is untrue although a detailed discussion is beyond the scope of this paper. Suffice it to say that both ATM and IP have their own relative strengths and weaknesses which explains why as they both evolve they do so towards a common point somewhere in between the two technologies as initially conceived. For example the introduction of resource reservations and per-flow state within IP networks in effect mimics the ATM philosophy while the introduction of the ABR(Available Bit Rate)[6] service into ATM provides a similar service to that of TCP over IP.

Regardless of technical comparisons between ATM and IP one observation cannot be ignored. Although ATM was conceived as the ubiquitious communications solution to take us into the next century it is now clear that such a role will instead be fulfilled by IP. In other words rather than considering IP as yet another protocol that can be carried

[1] This is one of the reasons why IP scales so well

over ATM it is perhaps more appropriate to consider ATM as yet another protocol that IP can operate over. However ATM is still destined to be a very important technology since, in the short term at least, commercial ATM switches will continue to offer a higher bit rate to cost ratio than IP routers. This benefit is complemented by ATM's support for QoS on demand while the short, fixed cell size facilitates low end-to-end delays. It is no surprise therefore that much interest has been generated with regard to how these desirable features of ATM can best be exploited in an IP internetwork.

Classical Approaches to IP over ATM

In some early schemes for IP over ATM deployment, leveraging the potential benefits of ATM has not been the primary concern. In fact the 'classical' models of 'Classical IP over ATM'[12] and 'LAN Emulation' [4] largely negate the potential benefits of ATM in exchange for maintaining the classical IP paradigm in order to facilitate easy migration to ATM. For example [12] separates an ATM network into Logical IP Subnets(LIS) interconnected by IP routers. Direct ATM connections between IP hosts in separate LISs are then prohibited even if the underlying ATM topology is capable of supporting them . Instead, an ATM VC originating within a given LIS can only extend as far as a router at the LIS boundary where the contents of the received ATM cells must then be reassembled into IP packets, each of which is then subjected to an IP forwarding decision and re-segmentation into ATM cells to be sent along the next intra-LIS VC along the journey as illustrated in Figure 1. The reassembly and re-segmentation at each router along the path severely restricts the potential benefits of ATM, namely those of a high bit rate, low latency path with QoS support. In addition because the end-to-end path is forced to traverse certain LIS boundary routers a convoluted path may result depending on the physical positioning of those boundary routers. Furthermore these classical models require additional address resolution mechanisms in order to map a next hop's IP address to its ATM address.

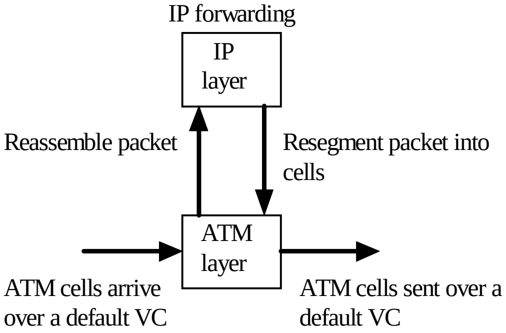
IP forwarding
IP layer
Reassemble packet
Resegment packet into cells
ATM layer
ATM cells arrive over a default VC
ATM cells sent over a default VC

Figure 1:Overhead of IP layer forwarding with classical models

In order to accommodate certain IP protocols, each LIS must provide intra-LIS broadcast which is typically implemented using a point-to-point VC from every node in the LIS to a multicast server and a single point-to-multipoint VC from the multicast server to every node in the LIS that it serves. This imposes a limit on the number of nodes in each LIS which is governed by the number of VCs that the multicast server can support. In addition there is a restriction on the overall size of the ATM network with this approach due to the necessity for an address resolution mechanism to map IP addresses to ATM addresses. The scalability of such an address resolution mechanism is particularly restricted by the fact that

ATM nodes within the same LIS need not be geographically contiguous which could make it difficult to use fully distributed database servers.

NHRP

The Next Hop Resolution Protocol (NHRP)[13] was developed as a means of facilitating inter-LIS VCs in order to utilise the potential benefits of ATM which are lost with the classical models. NHRP is an inter-LIS address resolution mechanism that maps a destination's IP address to the destination's ATM address in cases where the destination resides within the same ATM cloud as the source. In cases where the destination resides outside the ATM cloud containing the source, NHRP returns the ATM address of the source ATM cloud's egress router that is closest to the destination. Once the source receives the NHRP response it can then open a direct cut-through VC to the destination[2] using standard ATM signalling/routing protocols. However, opting to use NHRP and end-to-end VC setup for every single data flow in an NHRP capable network is unlikely to yield optimal results especially in large ATM clouds within the Internet. This is because in such an environment the number of IP flows traversing the cloud may be quite large in which case setting up a separate VC for each flow, may result in an unmanageable number of VCs[3] at switches within the cloud. Furthermore setting up a cut-through VC may be unnecessary and even undesirable for certain short-lived flows where it would be hard to justify the associated overhead[4] of the end-to-end connection and its setup, especially for flows that make no assumptions about QoS anyway. These issues are covered in more detail in [22] which suggests assigning responsibility for the VC cut-through decision to the sending application. Based on findings in [14] the majority of flows would not be suitable for VC cut-through and so would continue to be forwarded in a connectionless hop-by-hop manner over the default VCs as shown in Figure 2. However, the minority of flows that did receive cut-through would represent the majority of packets since these flows would be of much longer duration.

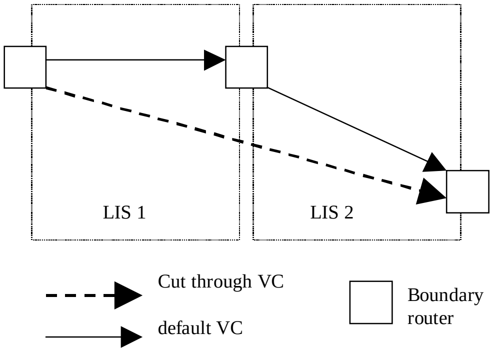

Figure 2:default "classical" service path vs cut-through service path

Although NHRP unquestionably overcomes some of the weaknesses of the classical IP over ATM models it is not without its own limitations. One of those is NHRP's inability to directly support multicast although certain elements of NHRP may be used to facilitate shortcuts

[2] or to the closest egress router to the destination in cases where the destination resides outside the ATM cloud

[3] A given ATM switch can only support a given number of VCs due to the limited VPI/VCI space. However perhaps a greater restriction is the rate at which VCs can be set up and torn down by the switch.

[4] Connection overhead comprises 3 main components. First, processing load of any control messages. Second, bandwidth consumed in sending any control messages. Third, usage of an additional VC.

within certain multicast scenarios such as [23]. Also an NHRP solution neccesitates routing/signalling functionality in both the ATM and IP layers which adds to the overall complexity.

IP Switch and Cell Switch Router

The proposals of IP switch[14] and Cell Switch router(CSR)[24] are based on similar hybrid ATM switch/IP router designs which allow coexistence of hop-by-hop IP forwarding with direct VC cut-through modes of service in order to provide each flow with the most suitable mode of service while maintaining desirable network conditions such as a manageable number of VCs. Although a number of similar schemes exist, notably IP on ATM[19] and IPSOFACTO[7], IP switch and CSR are the most well known and established, and as such are the focus of our discussion in this section. However one interesting difference is the fact that both IP switch and CSR rely on a signalling protocol to inform neighbouring nodes of any chosen flow-specific VCs whereas IP on ATM and IPSOFACTO inform the downstream node of any chosen flow-specific VCs implicitly through usage.

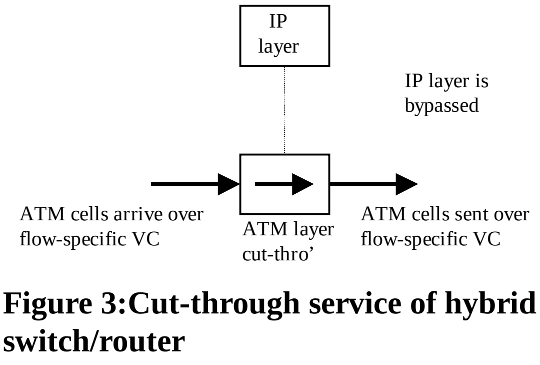

Figure 3:Cut-through service of hybrid switch/router

The IP switch and CSR hybrid switch/routers contain all the usual functionality of conventional IP routers and so are capable of providing a connectionless IP forwarding service as shown in Figure 1. However the valid topological configurations vary according to the type of hybrid switch/router. The IP switch does not support the ATM UNI standards[1]-[3] and so is incapable of interfacing with conventional ATM devices. Instead the IP switch would typically be used to replace each conventional ATM switch in existing ATM networks as shown in Figure 4. The CSR on the other hand is UNI-compatible and is therefore capable of interconnecting ATM-subnets in a similar way to the LIS border routers in the classical IP over ATM model, the difference being that unlike the classical models the CSR is also capable of providing direct VC cut-through between the adjacent subnets for selected data flows. Consequently valid configurations for CSRs include that of Figure 5 as well as that of Figure 4.

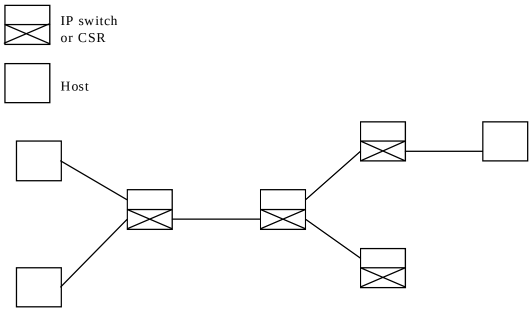

Figure 4: Valid topology for IP switch or CSR

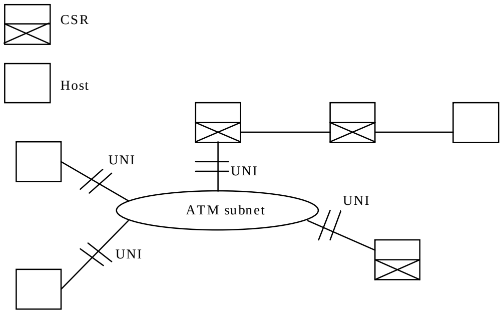

Figure 5:Valid topology for CSR

A dedicated cut-through VC for a specific flow can be implemented as shown in Figure 3 by associating an incoming VC with an outgoing VC in order to switch cells of that flow directly in hardware without IP forwarding. This cut-through service differs from that offered by NHRP and traditional ATM signalling in that the switching table associations are not made on an end-to-end basis. Instead each hybrid switch router makes a decision independently on whether or not to implement local cut-through. The local policy cut-through rules can be configured by network management and typically will result in cut-through for flows of any higher layer protocol that are suitable. For example TCP FTP flows are suitable since they are of sufficient duration to justify the overhead associated with cut-through setup. UDP(User Datagram Protocol) flows carrying NTP(Network Time Protocol) traffic on the other hand, where each flow typically consists of a single packet, are not suitable. The higher layer protocol can be determined by inspecting the packet headers during IP processing of the first packet of a flow. Once the cut-through decision has been made and the switching table associations installed, all further packets of the flow will receive local cut-through service. Once each hybrid switch/router along the end-to-end path has implemented local cut-through for a specific flow then the end-to-end service received by the flow is essentially the same as that obtained using the end-to-end signalling approach of NHRP. However, obtaining this service using a concatenation of local cut-throughs is advantageous in a number of respects. For example with the hybrid switch/routers the end-to-end route is determined entirely by the underlying IP routing protocols which means that the ATM routing/signalling and address resolution protocols are no longer required leading to a reduction in complexity[5]. In addition hybrid switch/ routers are well suited for use with RSVP which is the protocol of choice for setting up QoS over IP networks. RSVP control messages would travel over the default VCs and would receive full IP processing at each hybrid switch/router where they could initiate setup of flow-specific VC cut-throughs according to the QoS information contained within the RSVP messages. Also any VC associations that are setup by the hybrid switch/routers are soft-state which means that they need to be continually refreshed in order to avoid timeout. The use of soft-state rather than hard-state helps to maintain much of the connectionless nature of IP and is the same technique used by RSVP to good effect. Another key advantage of hybrid switch/routers compared to NHRP is that they offer full support for cut-through multicast trees by accommodating branch points at the ATM layer[6].

[5] However, CSRs must still support ATM UNI signalling in order to connect to adjacent CSRs that are reachable across ATM subnets.

[6] Although NHRP is a point-to-point mechanism, it could be used to emulate multicast through a number of point-to-point VCs although this will be bandwidth inefficient since many of the VCs

ARIS and Tag Switching

Aggregated Route Based IP Switching(ARIS)[26] and Cisco's Tag Switching architecture[21] are approaches to IP over ATM in which VC association is completely topology-driven unlike the hybrid switch/router models discussed in the previous section where setup of VCs is either topology driven for default-VCs or traffic-driven for flow-specific VCs[7].

Unlike the hybrid switch/router approaches both ARIS and Tag switching use VC cut-through for all traffic including best-effort. ARIS and Tag switching are able to do this without causing 'VC-explosion'[3] since the cut through VCs of both ARIS and Tag switching can have a courser granularity than the per-flow cut-through VCs of the hybrid switch/routers. In fact both ARIS and Tag switching offer a choice of granularities according to the network environment.

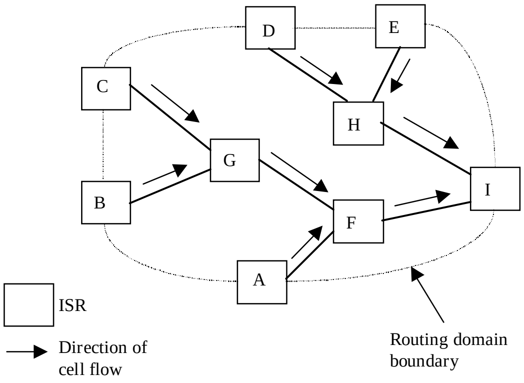

Figure 6: ARIS multipoint-pt tree

ARIS introduces the concept of "egress identifier" type to define granularity. For each value of "egress identifier" the ARIS protocol establishes a multipoint-to-point tree[8] that originates at routing domain ingress ISRs(Integrated Switch Router)[9] and terminates at the routing domain egress ISR[10] for that particular egress identifier. So if the egress identifier represents IP destination prefixes then a separate multipoint-to-point tree is set up per IP destination prefix. This is exemplified in Figure 6 which shows a multipoint-to-point tree that is set up from ingress ISRs A, B, C, D, E to egress ISR I. When a packet arrives at one of these ingress ISRs the forwarding table is consulted to determine the outgoing interface and VPI/VCI label to be used. Cells from the packet are then switched along the tree completely at the ATM layer until they reach the egress ISR I where the datagram is again reassembled at the IP layer.

The ARIS protocol mechanisms for setting up the tree vary depending on whether or not VC-merging is used.

VC merging is when cells arriving on separate incoming links of an ISR are routed onto the same VC of an outgoing link of the ISR. With AAL5 which has no intra-VC multiplexing identifier, VC merging is only possible provided no interleaving of cells from different AAL5 frames occurs. Otherwise it is not possible to reconstruct each AAL5 frame at the destination as there is no simple way of determining which cells belong to which frames. Switches that support VC merging do so by buffering cells of each incoming AAL5 frame until the full frame

may share common links which consequently carry the same data more than once.

[7] In addition CSRs permit flow-specific Vcs to be pre-configured.

[8] The multipoint-to-point tree will also be a multipoint-to-point VC if VC merging is used as described later in the section.

[9] This is the name that the ARIS specificiations use to refer to an ARIS-compatible switch.

[10] The identification of the egress ISR for a particular egress identifier is obtained from the routing protocols

has arrived, storing the full frame for a period of time determined by the scheduler, and then transmitting the frame so that it occupies a contiguous sequence of cells on the output link as shown in Figure 7. VC merging reduces the number of consumed VCs but adds latency due to buffering of AAL5 frames. However this increase in latency will still be less than for the case of IP forwarding while the switching speed will be close to that attainable without VC merging.

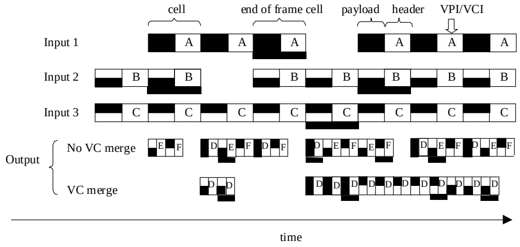

Figure 7:Ordering of output cells with and without VC merging.

If VC merging is not used by the ISRs then buffering of AAL5 frames is unnecessary since mapping each input VC to a separate output VC allows cells of frames from different input VCs to be interleaved on the output link while still being able to reconstruct each frame at the destination. This interleaving process is illustrated in Figure 7.

Tag switching uses a Tag Information Base(TIB) in each Tag Switch router in order to provide the mapping between an incoming interface and tag(VPI/VCI value) of an incoming cell to the outgoing interface and outgoing tag of the cell. The TIB entries can be installed either explicitly or using the Tag Distribution Protocol[10]. In the latter case a separate TIB entry is created for each route in the Forwarding Information Base(FIB)[11]. In addition the FIB is extended to include a tag entry for each route. Then when a packet first arrives at the ingress TSR for the tag switching network the FIB forwards the packet to the next hop while labelling the outgoing cells with the indicated tag value. Thereafter each TSR will switch the cells directly at the ATM layer using the TIBs of each subsequent TSR traversed.

The ARIS and Tag switching mechanisms are similar in other respects apart from those already mentioned. For example both mechanisms provide support for multicast and explicit routes. In addition both use default VCs between the hybrid switch/routers in order to implement hop-by-hop forwarding for their control protocols as well as for the IP routing protocols. Furthermore the ARIS and Tag switching architectures include protocol mechanisms to prevent the setup of switched path loops. Another common feature between the two mechanisms is that they are both able to correctly implement TTL decrement for cut-throughs. In other words when a packet is reassembled at the egress router following VC cut-through its TTL value will be the same as if it had undergone hop-by-hop IP forwarding instead. Apart from the throughput improvement obtained by ARIS and Tag switching through bypassing the IP layer the use of underlying ATM technology also makes them very suitable for offering QoS support although to date Tag switching has made more progress in this respect**[27**]

[11] The FIB is the information base in IP routers that is used to forward IP packets.

Discussion

The schemes that we have discussed regarding the integration of IP routing and ATM switching are currently receiving much attention within the networking community and this has resulted in the IETF setting up the Multiprotocol Label Switching(mpls)[12] Working Group in order to standardise these schemes. We can summarise the motivation behind these schemes as follows.

1) IP is the universal communications protocol.
2) ATM switches currently outperform IP routers.

If IP routers could handle the same bit rates as ATM switches without costing any more money then the benefits of deploying ATM beneath IP would be questionable. Although it has generally been regarded that ATM switches would always be faster than IP routers it appears likely that such a disparity will lessen in the future as we see more research effort directed towards high speed IP routers. Even at this early stage some key points can be noted as follows:

1) IP routers are being developed that use ATM cut-through internally on a per-packet basis rather than operating using the store and forward mechanism of traditional IP routers. In such devices ATM is shielded from the network in that ATM cells are never actually seen on the communications links. This means that as far as the network is concerned the device is an IP router.
2) IP routers are being developed that make use of novel techniques to achieve speeds of Gigabits/s and beyond. An example of one such commercial implementation that can achieve very high speeds is that of Pluris Inc[20].
3) Because IP packets can be much larger than ATM cells, less of them need to be processed per unit time to achieve the same bit rate. In this respect, IP is more suited to a higher bit rate than ATM.

Summary

In this paper we have looked at the principles upon which the Internet was developed and which have made it so successful. We have seen how a changing communications environment, particularly with regard to QoS and bandwidth demands, necessitates further evolution of the Internet if it is to maintain its position as the universal communications solution. ATM has many desirable characteristics such as QoS support as well as the ability to provide high bit rate, low latency end-to-end paths that are potentially useful within this context. We have examined the relative merits of various techniques for using ATM below an IP network. These include the use of hybrid (or integrated) switch/routers that do not adhere to the conventional design approach of strict separation between network layers. This is a fine example of ‘integrated layer processing’[9] in which carefully designed blurring of the boundaries between the IP and ATM layers allows them to be mutually supportive providing a solution that combines the speed of ATM with the flexibility of IP.

[12] The mpls working group of the IETF is concerned with label switching in general and not just the special case of label switching in an ATM environment although this is undoubtedly the major focus of attention.

Acknowledgements

The author would like to acknowledge British Telecom. Labs., Martlesham, England for supporting this work.

References

[1]ATM Forum (1993). ATM User Network Interface (UNI) Specification Version 3.0. AF-UNI-0010.001.

[2] ATM Forum (1994). ATM User Network Interface (UNI) Specification Version 3.1, AF-UNI-0010.002.

[3] ATM Forum (1996). ATM User Network Interface (UNI) Specification Version 4.0. AF-UNI-4.0.

[4]ATM Forum. LANE Client Management Specification Version 1, September 1995 . ftp://ftp.atmforum.com/pub/approved-specs/af-lane-0038.000.ps

[5]ATM Forum. LANE Servers Management Specification Version 1, March 1996 .

[6] ITU. Recommendation I.371 (08/96) - Traffic control and congestion control in B-ISDN

[7]A. Acharya, R. Dighe, F. Ansari. IPSOFACTO: IP Switching Over Fast ATM Cell Transport, Internet Draft, July 1997, draft-acharya-ipsw-fast-cell-00.txt.

[8]D. Cansever. NHRP Protocol Applicability Statement, Internet Draft, February 1996, draft-ietf-rolc-nhrp-appl-02.txt.

[9] D. Clark and D. Tennenhouse. Architectural Considerations for a New Generation of Protocols. ACM SIGCOMM, pages 200-208, September 1990, Philadelphia.

[10] P. Doolan et al. Tag Distribution Protocol, Internet Draft, May 1997, draft-doolan-tdp-spec-01.txt.

[11] N. Feldman and A. Viswanathan. ARIS protocol specification, Internet Draft, March 1997, draft-feldman-aris-spec-00.txt.

[12] M. Laubach. Classical IP and ARP over ATM. Request for Comments, January 1994, RFC1577.

[13] J. Luciani et al. NBMA Next Hop Resolution Protocol (NHRP), Internet Draft, draft-ietf-rolc-nhrp-13.txt

[14]P. Newman, T. Lyon and G. Minshall. Flow Labelled IP: A Connectionless Approach to ATM, Proceedings of IEEE INFOCOM, San Fransisco, March 1996.

[15]P. Newman et al. IP Switching and Gigabit Routers, IEEE Communications magazine, Jan 1997.

[16] P. Newman et al. Transmission of Flow Labelled IPv4 on ATM Data Links Ipsilon Version 1.0, Request for Comments, May 1996: RFC1954.

[17] P. Newman et al. Ipsilon Flow Management Protocol specification for IPv4 Version 1.0, Request For Comments, May 1996, RFC1953.

[18] IETF Multiprotocol Label Switching(mpls) Working Group,

http://www.ietf.org/html.charters/mpls-charter.html

[19]G. Parulkar, D. Schmidt and J. Turner. A Strategy for Integrating IP with ATM, SIGCOMM'95, ftp://ftp.atm.atmforum.com/pub/approved-specs/af-lane-0057.000.ps

[20] Pluris Inc, Terabit Routing Company web site, http://www.pluris.com

[21] Y. Rekhter, B. Davie, D. Katz, E. Rosen, G. Swallow. Cisco System' s Tag Switching Architecture Overview, Request for Comments, Feb 1997, RFC2105.txt

[22] Y. Rekhter and D. Kandlur. Local/Remote Forwarding Decision in Switched Data Link Subnetworks, Request for Comments, May 1996, RFC1937.

[23]Y. Rekhter and D. Farinacci. Support for Sparse Mode PIM over ATM, Internet Draft, February 1996, draft-rekhter-pim-atm-00.txt.

[24] Toshiba Corporation. Cell Switch Router white paper Version 1.0, November 1996.

[25]P. White and J. Crowcroft. The Integrated Services in the Internet: State of the Art, Proceedings of IEEE, December 1997, Volume 85, No. 12, pp 1934-1946. ftp://cs.ucl.ac.uk/darpa/rsvp4.ps

[26] R. Woundy, A. Viswanathan, N. Feldman, R. Boivie. ARIS: Aggregate Route-Based IP Switching, Internet Draft, Nov 1996, draft-woundy-ais-ipswitching-00.txt

[27] A. Lin, B. Davie, F. Baker. Tag Switching Support for Classes of Service, Internet Draft, Dec 1996, draft-lin-tags-cos-00.txt