5. Summaries

A new internetworking architecture based on the Cell Switch Router (CSR) was described. Possible use and protocol overview of the CSR, and migration scenario from conventional ATM internetworking architecture based on the ATM Forum's LAN Emulation or IETF's RFC1577 to CSR-based internetworking which would cooperate with the MPOA-based architecture was investigated. It provides users with high-throughput and low-latency ATM pipes while keeping conventional "router-based" internetworking strategy as they are. It also provides users with specific QoS/bandwidth by cooperating with internetworking level resource reservation protocols such as RSVP.

We have been implementing an experimental CSR system using an ATM switch (cell switching function of CSR) and a PC (IP routing, datagram forwarding, and Bypass-pipe control functions of CSR) which are interconnected by a 155Mbps ATM line and a serial line for ATM switch control by the PC. A detailed description of the experimental system and evaluation results will be presented shortly.

References

[MARS95] G. Armitage, "Support for Multicast over UNI 3.1 based ATM Networks", IETF Internet Draft (work in progress), draft-ietf-ipatm- ipmc-10.txt, Dec. 1995.

[IPNNI94] R. Callon, "Integrated PNNI for Multi-Protocol Routing", The ATM Forum Contribution No. 94-0789, Sept. 1994.

[NHRP95] D. Katz and D. Piscitello, "NBMA Next Hop Resolution Protocol(NHRP)", IETF Internet Draft (work in progress), draft-ietf- rolc-nhrp-06.txt, Nov. 1995.

[LANE95] The ATM Forum, "LAN Emulation over ATM Version 1.0", Jan. 1995

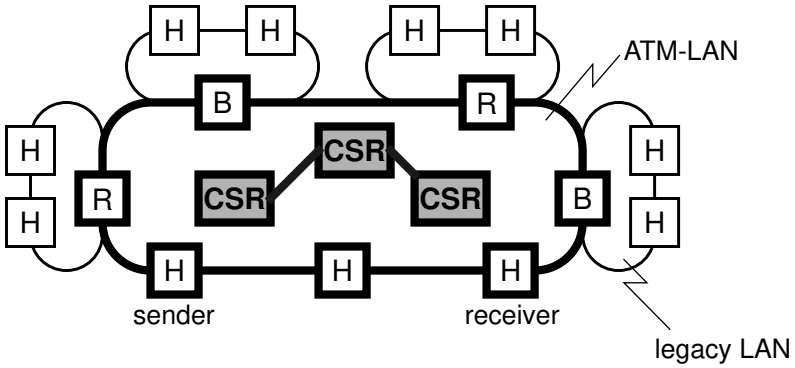
[PNNI95] The ATM-Forum, "P-NNI Draft Specification R15", Nov. 1995.

[REK95] Y. Rekhter, et al., "Local / Remote Forwarding Decision in Switched Data Link Subnetworks", IETF Internet Draft (work in progress), draft-ietf-rolc-apr-03.txt, Nov. 1995.

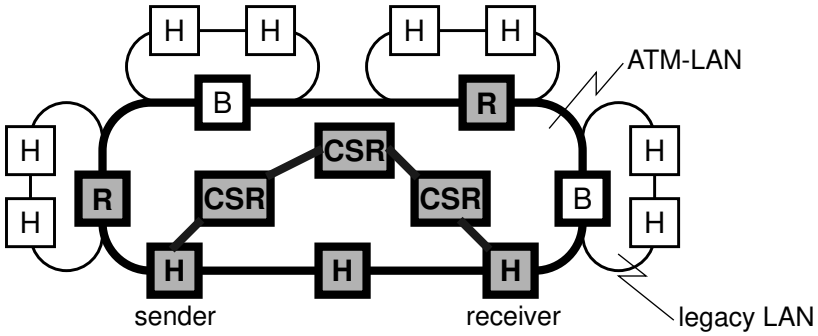
[RFC1577] M. Laubach, "Classical IP and ARP over ATM", IETF RFC 1577, Oct. 1993.

[RSVP95] R. Braden, et al., "Resource ReSerVation Protocol (RSVP), Version 1 Functional Specification", IETF Internet Draft (work in progress), draft-ietf-rsvp-spec-08.ps/txt, Nov. 1995.

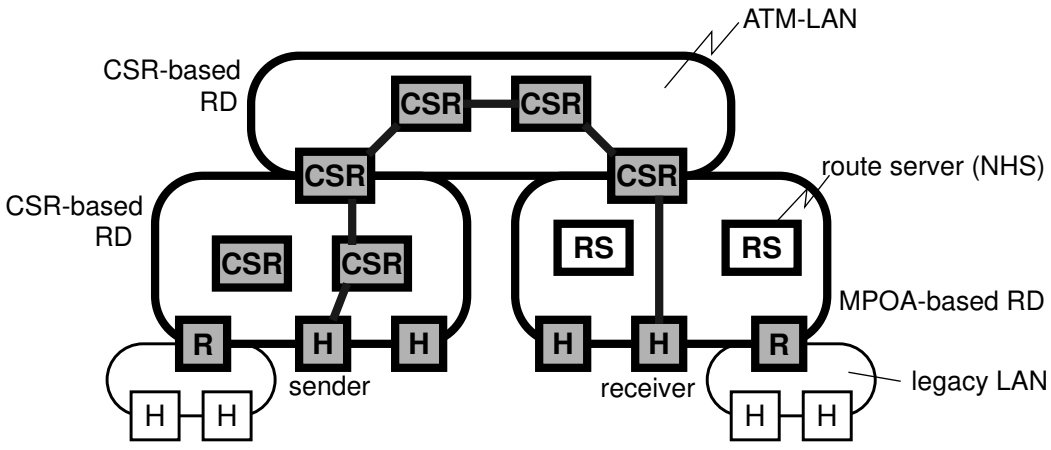
[STII95] L. Delgrossi and L. Berger, "Internet STream Protocol Version 2(STII)", Internet Draft (work in progress), draft-ietf-st2- spec-04.txt, Aug. 1995.



Step 1 : CSR to CSR Bypass-pipe



Step 2 : Host/EdgeRouter to Host/EdgeRouter Bypass-pipe



Step 3 : Global Bypass-pipe in cooperation with MPOA

Figure 4 Three Step Migration Scenario of Cell Switch Router

any changes in conventional ATM devices such as routers, bridges or hosts, high-throughput path or QoS-ed paths can be provided only from the first-hop CSR to the last-hop CSR. Datagram transmission from/to conventional ATM devices to/from CSRs would be based on either the ATM Forum's LAN Emulation [LANE95] or the RFC1577 Classical IP Model in this step.

Step 2 : Provision of host/edge router to host/edge router Bypass-pipe

At this stage of CSR migration, software functionalities to become edges of Bypass-pipes will be implemented in conventional ATM devices such as ATM routers or ATM hosts as well as CSRs. High-throughput paths or QoS-ed paths can be provided from ATM host/router to ATM host/router via CSRs in this step. Since ATM bridges do not analyze IP level information or information higher than that, they cannot become edges of Bypass-pipes without any modification. Datagram transmission from/to ATM bridges would be based on LAN Emulation, while any intra-subnet datagram transmission from ATM host/router to ATM host/router would still be based on LAN Emulation or RFC1577.

Step 3 : Provision of global Bypass-pipe in cooperation with MPOA

The MPOA (Multiprotocol over ATM) specification will be completed by early 1997 by the ATM Forum. It enables direct ATM-VC setup from ATM host/edge device to ATM host/edge device regardless of internetworking addresses of its source and destination. Since MPOA will mostly be applied just within the same routing domain (RD) in order to avoid routing loop creation, MPOA-based direct VCs will be terminated at RD boundary routers. The CSR can be utilized as an RD boundary router that concatenates MPOA-based VCs to create inter-RD Bypass-pipes, although the CSR can still be utilized as an intra-RD router. ATM hosts/edge devices in this case should be equipped with Bypass-pipe edge functionalities as well as MPOA client functionalities.

transmitted from the sender hosts or routers at the upstream edge toward receiver hosts or routers at the downstream edge. In order to avoid every router on the path to initiating Bypass-pipe setup procedures for the same IP flow, trigger detection functions would be limited to sender hosts or routers at the upstream edge (ingress points of an ATM cloud).

Since the purpose of the Bypass-pipe setup in this case is not provision of a specific bandwidth/QoS but provision of direct ATM connectivity, heterogeneity of receivers does not have to be taken into account, this makes sender-initiated control much easier than the receiver-initiated one. Bypass-pipes will be set up towards receiver hosts or routers at the downstream edge (egress points of an ATM cloud). In order to support dynamic Bypass-pipe control capabilities which follow IP routing changes and to provide multicast Bypass-pipes as well, a soft-state protocol that transmit setup/refresh messages periodically is adopted.

b) Protocol for long-term use

In the case that Bypass-pipe setup is initiated as a result of specific bandwidth/QoS request from receiver hosts with RSVP, the setup will be carried out from receiver-side to sender-side. As RSVP itself does not have capabilities to exchange mapping information between reserved IP flows and ATM connections, information in addition to RSVP *Resv* messages need to be exchanged between adjacent routers (or hosts). This will be done by transmitting additional messages in cooperation with *Resv* messages. The route of the Bypass-pipe follows the IP routing information since RSVP *Path* messages determine the path of datagrams based on the IP routing information.

Although heterogeneity of receivers is taken into account in RSVP, how to allocate ATM connections in such heterogeneous and dynamically chageable environments with regard to bandwidth/QoS requires further study.

4. Migration Scenario of CSR-based Internetworking

A three step migration scenario for the CSR-based internetworking architecture is described as follows (see figure 4).

Step 1 : Provision of CSR to CSR Bypass-pipes only

At the primary stage of CSR migration, only CSRs, which has cell-switching capability, have software functionalities to become edges of Bypass-pipes (handling of Bypass-pipe messages and emission of datagrams to Bypass-pipes) as well as intermediate points of them. Although this step does not require

case, the purpose of Bypass-pipe setup in each of these cases is to reduce IP processing burden at intermediate routers as well as to provide a communication path with low latency for burst data transfer, rather than to provide end host applications with specific bandwidth/QoS. There would be no rule for determining bandwidth for such kinds of Bypass-pipes since no explicit information about bandwidth/QoS requirement is available without IP-level bandwidth/QoS management protocols such as RSVP. Using UBR VCs as components of the Bypass-pipe would be the easiest choice although there is no guarantees for cell loss quality, while using other services such as CBR/VBR/ABR with an adequate parameter tuning would be possible.

b) Provision of specific bandwidth/QoS requested by end hosts *(long-term use)*

This indicates cases in which routers or end hosts initiate a Bypass-pipe setup procedure by triggers related to IP-level bandwidth/QoS request from end hosts. The "Resource management entity" in the host or router, which has received bandwidth/QoS requests from applications or adjacent nodes may choose to accommodate the requested IP flow (identified by the source/destination IP address and port number for instance) to an existing VC or choose to allocate a new Dedicated-VC for the requested IP flow. Selecting the latter choice at each router can correspond to the trigger for constituting a Bypass-pipe. When both an incoming VC and an outgoing VC are dedicated to the same IP flow(s), those VCs can be concatenated at the CSR (ATM cut-through) to constitute a Bypass-pipe. Bandwidth for the Bypass-pipe (namely, individual VCs constituting the Bypass-pipe) in this case would be determined based on the bandwidth/QoS requirements by the end host which is conveyed by, e.g., RSVP messages. The ATM service classes; e.g., CBR/VBR/ABR; that would be selected depends on the IP level service classes requested by the end hosts.

Bypass-pipe provision for the purpose of b) will surely be beneficial in the future when related IP-level resource reservation protocol (e.g., RSVP) will become available as well as when definitions of individual service classes and flow specs offered to applications become clear. On the other hand, Bypass-pipe provision for the purpose of a) may be beneficial right now since it does not require availability of IP-level resource reservation protocols. In that sense, a) can be regarded as a kind of short- term use while b) is a long-term use.

3.2 Protocol Overview for CSR

a) Protocol for short-term use

In the case that a Bypass-pipe setup is initiated when hosts or routers detect datagrams with certain higher layer information (e.g, detection of TCP SYN-flag for ftp, etc.), Bypass-pipe setup requests are

services over the ATM cloud. And realization of IP level resource reservation protocols such as RSVP in NHRP environments requires further investigation.

The CSR-based internetworking architecture which keeps subnet-by-subnet internetworking with regard to any control protocol sequence does not require any extensions in multicast ATMARP techniques described in [MARS95] nor does it require any change in RSVP control procedure [RSVP95]. An example of a Bypass-pipe configuration for IP multicast is shown in figure 3. ATM point-to-multipoint VCs are established for individual subnet level multicast, which are concatenated at individual CSRs to configure global end-to-end IP multicast based on a Bypass-pipe.

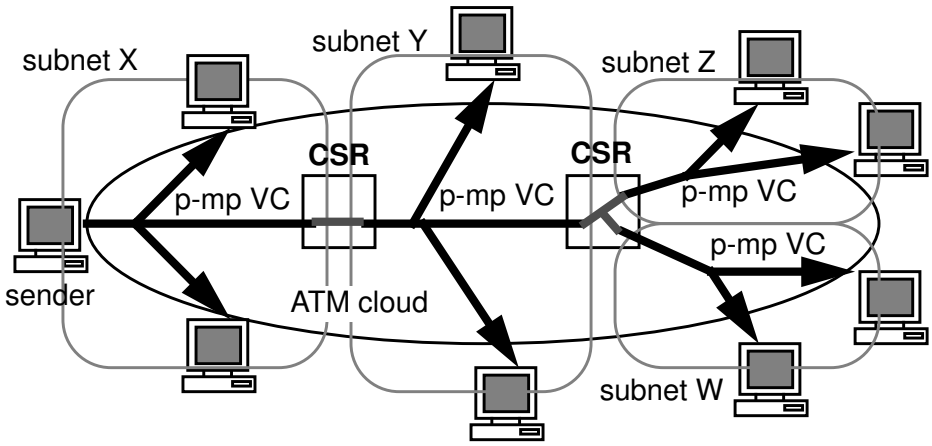


Figure 3 CSR-based IP Multicast Cut-through over ATM

3. Possible Use and Protocol Overview of CSR

3.1 Possible Use of Bypass-pipe

Possible use (or purposes) of Bypass-pipes provided by the CSR, in other words, possible triggers that initiate Bypass-pipe setup procedure, is discussed in this subsection.

a) Provision of low latency path *(short-term use)*

This indicates cases in which end hosts or routers initiate a Bypass-pipe setup procedure when they will transmit large amount of datagrams. Triggers for such kind of Bypass-pipe setup initiation will be, for instance, detection of datagrams with certain higher layer information such as TCP SYN-flag for ftp, nntp and http. Other triggers may be possible depending on the policy in each network. In any

2.2 Features

The main feature of the proposed CSR-based internetworking architecture is the same as that of the NHRP-based architecture in the sense that they both provide direct ATM level connectivity beyond IP subnet boundaries. There are, however, several notable differences in the CSR-based architecture compared to the NHRP-based one as follows.

1) Relationship between IP routing and ATM routing

In the NHRP model, after an egress point of the ATM cloud is first resolved based on IP level routing information, the actual route for an ATM-VC to the obtained egress point is determined based on ATM level routing information. We need to manage two different routing protocols over the large ATM cloud until Integrated-PNNI [IPNNI94] (which takes both ATM level metric and IP level metric into account) will be phased in in the future.

In the CSR model, IP level routing determines an inter-subnet level paths that shows which CSRs it should pass through. ATM level routing determines an intra-subnet level path for ATM-VCs (both Dedicated-VC and Default-VC) only between adjacent nodes (CSRs or ATM-attached hosts/routers). Since the roles of routing are hierarchically subdivided into inter-subnet level (router level) and intra-subnet level (ATM switch level), ATM routing does not necessarily have to operate all over the ATM cloud but only in individual IP subnets independent from each other. This will decrease the amount of information for complex ATM routing protocol handling. But an end-to-end ATM path may not be optimal compared with the NHRP model since the path should go through routers at subnet boundaries in the CSR model.

2) Dynamic routing and redundancy support

A CSR-based network can dynamically change routes for Bypass-pipes when related IP level routing information changes. Bypass-pipes related to the routing changes do not have to be torn down nor established from scratch since intermediate CSRs related to IP routing changes can follow them and change routes for related Bypass-pipes by themselves. The same things apply when some error or outage happens in any ATM nodes/links/routers on the route of a Bypass-pipe. CSRs that have noticed such errors or outages would change routes for related Bypass-pipes by themselves.

3) Interoperability with IP level resource reservation protocols in multicast environments

As current NHRP specification assumes application of NHRP to unicast environments only, multicast routers with multicast ATMARP client [MARS95] capability are still required to provide IP multicast

The route for a Bypass-pipe is determined when it is set up based on the IP routing table in each CSR. In figure 2, IP datagrams from source host or router X.1 to destination host or router Z.1 are transferred over the route X.1 -> CSR1 -> CSR2 -> Z.1 regardless of whether the communication is on a hop-by-hop basis or Bypass-pipe basis. Routes for individual Dedicated-VCs which constitutes the Bypass-pipe X.1 --> Z.1 (X.1 -> CSR1, CSR1 -> CSR2, CSR2 -> Z.1) would be determined based on ATM routing protocols such as PNNI [PNNI95], and would be independent of IP level routing.

An example of IP datagram transmission mechanism is as follows;

- The host/router X.1 checks an identifier of each IP datagram, which may be the “destination IP address (prefix)”, “source/destination IP address (prefix) pair”, “destination IP address and port”, “source IP address and Flow label (in IPv6)”, and so on. Based on either of those identifiers, it determines over which VC the datagram should be transmitted.

- The CSR1/2 checks the VPI/VCI value of each incoming cell. When the mapping from the incoming interface/VPI/VCI to outgoing interface/VPI/VCI is found in an ATM routing table, it is directly forwarded to the specified interface through an ATM switch module. When the mapping in not found in the ATM routing table (or the table shows an IP module as an output interface), the cell is assembled to an IP datagram and then forwarded to an appropriate outgoing interface/VPI/VCI based on an identifier of the datagram.

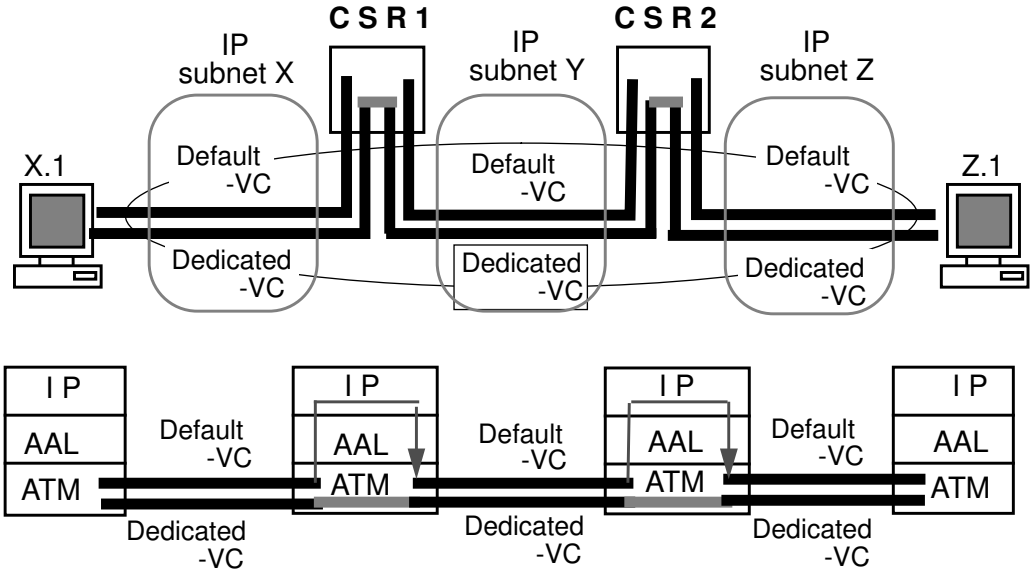


Figure 2 CSR-based Internetworking Model

2. Internetworking Architecture Based On the Cell Switch Router (CSR)

2.1 Overview

The Cell Switch Router (CSR) is a key network element of the proposed internetworking architecture. The CSR provides cell switching functionality in addition to conventional IP datagram forwarding. Communications with high throughput and low latency, that are native properties of ATM, become possible by using this cell switching function even when the communications pass through IP subnetwork boundaries. In an ATM internet composed of CSRs, VPI/VCI-based cell switching which bypasses datagram assembly/disassembly and IP header processing is possible at every CSR for communications which lend themselves to such (e.g., communications which require specific bandwidth and QoS, or which transmit large amount of data), while hop-by-hop datagram forwarding based on the IP header is also possible at every CSR for other conventional communications.

By using such cell-level switching capabilities, the CSR is able to concatenate incoming and outgoing ATM VCs, although the concatenation in this case is controlled outside the ATM cloud unlike conventional ATM switch nodes. That is, the CSR is attached to ATM networks via an ATM-UNI instead of NNI. By carrying out such VC concatenations at multiple CSRs consecutively, ATM level connectivity (which we call "ATM Bypass-pipe") composed of multiple ATM VCs, each of which connects adjacent CSRs (or CSRs and hosts/routers), can be provided.

Example network configurations based on CSRs are shown in figure 2. Two different kinds of ATM VCs are defined between adjacent CSRs or between CSRs and ATM-attached hosts/routers.

1) Default-VC : is a general purpose VC used by any communications which select conventional hop-by-hop IP routed path. All incoming cells received from this VC are assembled to IP datagrams and handled based on their IP headers.

2) Dedicated-VC : is used by specific communications (IP flows) which are specified by, for example, any combination of the destination IP address/port, the source IP address/port or IPv6 flow label. It can be concatenated with other Dedicated-VCs which accommodate the same IP flow as it, and can constitute an ATM Bypass-pipe for those IP flows.

Ingress/egress nodes of the Bypass-pipe can be either CSRs or ATM-attached routers/hosts both of which understand a Bypass-pipe control protocol. On the other hand, intermediate nodes of the Bypass-pipe should be CSRs since they need to have cell switching capabilities as well as understand the Bypass-pipe control protocol.

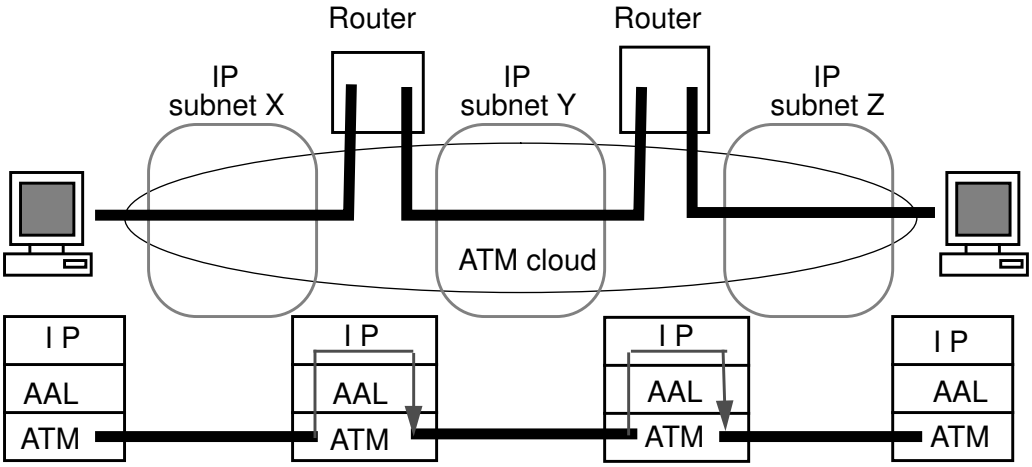
RSVP itself is not a connection oriented technology since datagrams can be transmitted regardless of the result of the resource reservation process. After a resource reservation process from a receiver(s) to a sender(s) is successfully completed, RSVP-capable routers along the path of the flow reserve their resources for datagram forwarding according to the requested flow spec. STII is regarded as a connection oriented IP which requires connection setup process from a sender to a receiver(s) before transmitting datagrams. STII-capable routers along the path of the requested connection reserve their resources for datagram forwarding according to the flow spec.

Neither RSVP nor STII restrict underlying datalink networks since their primary purpose is to let routers provide each IP flow with desired forwarding quality (by controlling their datagram scheduling rules). Since various datalink networks will coexist as well as ATM in the future, these IP level resource reservation technologies would be necessary in order to provide end-to-end IP flow with desired bandwidth and QoS.

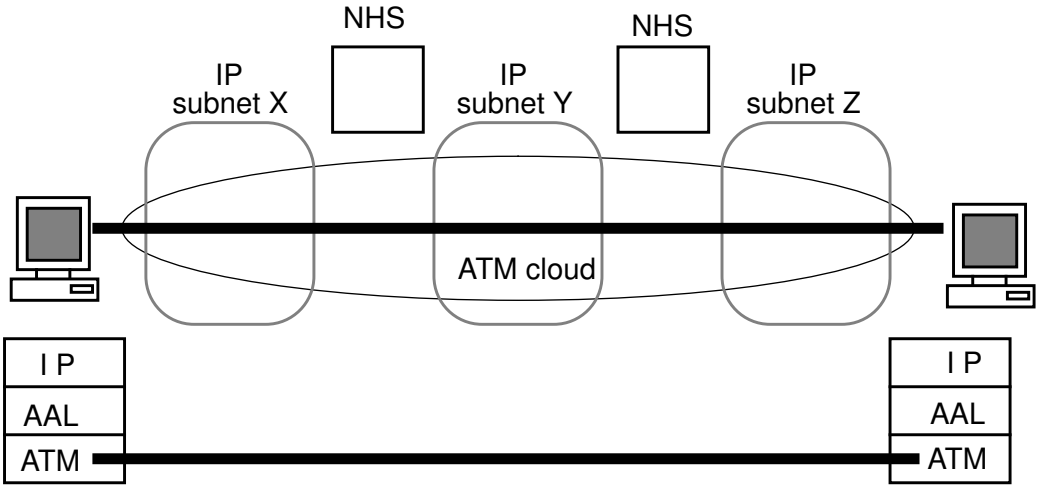
Taking this background into consideration, we should be aware of several issues which motivate our proposal.

- As of the time of writing, the ATM specific internetworking architecture proposed (as the NHRP model) does not take into account interoperability with IP level resource reservation or connection setup protocols. In particular, operating RSVP in the NHRP-based ATM cloud seems to require much effort since RSVP is a soft-state receiver-oriented protocol with multicast capability as a default, while ATM with NHRP is a hard-state sender-oriented protocol which does not support multicast yet.
- Although RSVP or STII-based routers will provide each IP flow with a desired bandwidth and QoS, they have some native throughput limitations due to the processor-based IP forwarding mechanism compared with the switching mechanism of ATM.

The main objective of our proposal is to resolve the above issues. The proposed internetworking architecture makes the best use of the property of ATM by extending legacy routers to handle future IP features such as flow support and resource reservation with the help of ATM's cell switching capabilities.



(a) Classical IP Model



(b) NHRP Model

Figure 1 Classical IP Model and NHRP Model

2) IP level resource reservation and flow support

Apart from investigation on specific high-speed datalink technology such as ATM, resource reservation technologies in router networks for desired IP level flows have been studied. Their typical examples are RSVP [RSVP95] and STII [STII95].

ever they go beyond IP subnet boundaries even though their source and destination are accommodated in the same ATM cloud. Although this model does not require drastic changes to the legacy internetworking architectures with regard to the IP datagram forwarding process, it still has problems of limited throughput and large latency compared with the ability of ATM, due to IP header processing at every router. It will become more critical when multimedia applications that require much larger bandwidth and smaller latency become dominant in the near future.

Another internetworking model is currently under discussion in the IETF ROLC WG and the ATM Forum MPOA (Multiprotocol-over-ATM) SWG. The model, that we call "NHRP (Next Hop Resolution Protocol) Model" here, aims at resolving throughput and latency problems in the Classical IP Model and making the best use of ATM. The Next Hop Server [NHRP95] (or Route Server) is introduced which can find an egress point of the ATM cloud nearest to the given destination and resolve its ATM address. After the ATM address of a desired egress node is resolved, the ingress node establishes a direct ATM connection to the obtained egress node even when they do not share the same IP address prefix. Once a direct ATM connection has been set up through this procedure, IP datagrams do not have to experience hop-by-hop IP processing but can be transmitted over the direct ATM connection. Therefore, high throughput and low latency communications become possible even if they go beyond IP subnet boundaries. In this model, ATM is utilized not only as a datalink function but also as a replacement of current hop-by-hop IP forwarding function. However, router-based hop-by-hop IP datagram forwarding functionality would still be required in the following cases;

- When you want to transmit IP datagrams before direct ATM connection from an ingress point to an egress point of the ATM cloud is established
- When you neither require a certain QoS nor transmit large amount of IP datagrams for some communication [REK95]
- When the direct ATM connection is not allowed for security or policy reasons

Cell Switch Router
- Basic Concept and Migration Scenario -

Yasuhiro Katsube Ken-ichi Nagami Hiroshi Esaki
(katsube, nagami, hiroshi@isl.rdc.toshiba.co.jp)

R&D Center, Toshiba Corporation

Abstract

A new internetworking architecture based on Cell Switch Router (CSR), which has ATM cell switching capabilities in addition to conventional datagram forwarding, is described. Possible use and protocol overview of CSR, and migration scenario from conventional ATM internetworking architecture based on LAN Emulation or RFC1577 to CSR-based internetworking which finally would cooperate with the ATM Forum's MPOA-based architecture is investigated. It provides users with high-throughput and small-latency ATM pipes while keeping conventional "router-based" internetworking strategy as they are. It also provides users with specific QoS/bandwidth by cooperating with internetworking level resource reservation protocols such as RSVP.

1. Background and Motivation

The Internet is growing both in its size and its traffic volume. In addition, recent applications often require guaranteed bandwidth and Quality of Service (QoS) rather than best-effort. Such changes are about to make the current hop-by-hop best-effort datagram forwarding paradigm inadequate, and accelerate investigations on new internetworking architectures.

Roughly two major schools of approaches can be seen as possible solutions whose main purpose is to improve ability of the Internet with regard to its throughput and QoS. One is to utilize ATM technology as much as possible, and the other is to introduce the concept of resource reservation and "flow" into IP.

1) Utilization of ATM

A typical internetworking architecture specified by the IETF IPoverATM Working Group (WG) is the "Classical IP Model" [RFC1577]. As shown in figure 1, this model allows direct ATM connectivities only between nodes that share the same IP address prefix. IP datagrams should traverse routers when-