PAPER

High Speed Datagram Delivery over Internet using ATM Technology

Hiroshi ESAKI†, Masataka OHTA†† *and* Ken-ichi NAGAMI †, *Members*

SUMMARY This paper proposes a high throughput small latent IP packet delivery architecture using ATM technology in a large scaled internet. Data-link network segments, including ATM network segments, are interconnected through routers.

A connection oriented IP packet delivery will be provided by IP (including both IPv4 and IPv6) with a certain resource reservation protocol (e.g. RSVP). When the router attached to ATM network segment has a mapping function between the flow-ID (e.g.in the SIPP header) and the VPI/VCI value, the small latent connection oriented IP forwarding can be provided.

Also, when the router has cell-relaying functionality, the small latent connectionless IP forwarding can be provided, even in IPv4. The source router, where the source end-station belongs to, will be able to transfer the connectionless IP packet to the destination router, where the destination end-station belongs to, through the concatenated ATM connections (ATM-VCCs) without any ATM-VCC termination point. When all of the network segments are ATM-LAN, the proposed architecture can accommodate about up to $2^{22} (\simeq 4 \times 10^6)$ end-stations with two network layer processing points . And when the network is scaled up hierarchally, we can accommodate larger number of end-stations. For example, we can accommodate 10^{15} end-stations by a three layered network. Then the maximum number of actual network layer processing points between source and destination end-stations can be ten. Here, 10^{15} is the maximum number of end-stations in ISDN and also it is the target number of accommodated end-stations for IPv6.

key words: ATM, Internet, IP, bypass, cell-relaying, re-direction

1. Introduction

ATM (Asynchronous Transfer Mode) will be the solution of communication platform, providing a highly flexible and cost-effective network for handling a wide variety of communications. The user information is handled by the small sized packet, called cell [1][3]. Since the data format to be dealt with by an ATM network is independent of the user application, it is said that an ATM network could provide flexible communication pipes to the user applications, effectively. In an ATM network, all the information is transferred through a virtual connection (VCC/VPC), which is a connection oriented virtual pipe between the source and destination end-points. Through the high speed virtual connection, each end-station can transfer the information with small latency.

ATM is recognized as one of the platforms to provide a high speed QOS (Quality of Service) data-link layer service. Here, the QOS data-link layer service can be end-to-end, rather than link-by-link, in ATM networks. But, it should be still recognized as a data-link layer service, rather than a network layer service. An ATM network provides a connection oriented service as well as a connectionless service (e.g. conventional IP forwarding) topping on connection oriented VCCs, which could have a certain QOS.

Data-link network segments, including ATM-LANs, should be interconnected through routers (network layer entities), even when the ATM becomes a major data-link platform. From Internet's point of view, data-link segments should be interconnected through routers, and they should be operated independently [11]. The network segment (autonomous system) using ATM technology will increase in the future. However, ATM-LANs are interconnected with ATM-LANs, as well as other network platforms (e.g.high speed Ethernet). And, old network platforms must be smoothly replaced by ATM technology or by the other high speed platforms. Also, from the view point of network operation, each network segment should be rationally small, and should be operated independently.

IETF (Internet Engineering Task Force) has discussed IPng (IP next generation : IPv6), which will replace the current IP (i.e. IPv4). IPv6 is determined as SIPP16 [7] in 30th IETF meeting. SIPP16 has a 128 bit length address field and does have a 28 bit length flow-ID field. Flow-ID is a unique value for every source destination 128 bit address pair. Therefore, multiple packet flows can be defined for every source destination address pair. In the IPv6 environment, both connection oriented and connectionless IP forwarding will be provided. In order to provide connection oriented IP forwarding, some kind of protocol (e.g. RSVP : Resource ReSerVation Protocol) will be applied [10].

This paper proposes a high speed and small latent IP packet delivery architecture over the Internet, using ATM technology. Regarding a connectionless service, we can provide a small latent IP packet delivery using a cell-relaying bypass route, not only for IPv6

Manuscript received August 23, 1994.
Manuscript revised March 1, 1995.

†The authors are with RCenter, Toshiba Corporation, Kawasaki-shi, 210 Japan.
††The author is with Tokyo Institute of Technology, Meguro-ku, 152 Japan.

networks but also for IPv4 networks. Regarding a connection oriented (network layer) service, we can provide a high throughput small latent IP packet delivery using router's mapping functionality between flow-ID in the IP packet header and VPI/VCI value. Also, when router has a mapping capability between the network layer QOS parameter (flow-spec) and the ATM's QOS parameter, an end-to-end network layer QOS will be provided.

Structure of the rest of this paper is as followed. In section 2, the general network architecture of the Internet, including ATM-LANs, is discussed. In section 3, the architecture of ATM router to provide connection oriented IP forwarding is proposed. In section 4, the architecture of ATM router to provide a small latent connectionless IP packet delivery is discussed and evaluated in the ideal case which is the case where all of data-link segments are based on ATM technology. And, section 5 gives a brief summary. Here, "routing" in this paper means routing at network layer (or internetwork layer) : i.e. data-link layer routing is not referred as routing.

2. Internet Architecture including ATM-LANs

2.1 Network Architecture in ATM Forum and IETF

Models developed at LE-WG (LAN Emulation Working Group) of ATM Forum and at ATM-WG and ROLC(Routing Over Large Cloud)-WG of IETF are discussed [5][6].

Basically two architecture models are discussed. Let say them as CLIP-model and NBMA-model, in this paper.

- CLIP-model [5]
  Generally, (logical) IP subnets will overlay on a single ATM network segment, which may be a large scaled network. This means that end-stations, which generally belong to different IP subnets, will exist on a single (ATM-ed) data-link segment.
  IP subnets are interconnected by routers, which is completely the same as the conventional router. The router in [5] only has an IP forwarding capability, but it does not have a cell relaying capability. This means that ATM-VCC shall be always terminated at routers. Therefore, it is impossible to provide end-to-end QOS-ed virtual connections using CLIP-model, without some protocols discussed in IETF, e.g.[10]. Here, even when RSVP is applied in CLIP model, datagram transferring is still performed by IP packet level and it causes both delay and HOL (Head Of Line) blocking.

- NBMA-model [6]
  Logical IP subnets will overlay on a single ATM network segment. NBMA-subnet is defined as the network domain where an end-to-end ATM-VCC can be established.
  Within the NBMA-subnet, each end-station can establish (seamless) ATM-VCC toward another end-station that is located at the other IP subnet, without passing through a router. This means that the current security function (e.g. fire-wall) does not work, and that a new security function must be developed.
  When the IP packet should be transferred to outside the NBMA-subnet, it will be transferred to the border router of NBMA-subnet. This means that, even if the neighbor NBMA-subnet is an ATM network, it is impossible to establish an end-to-end ATM-VCC using IP addresses.

2.2 Internet Architecture including ATM Platforms

To operate a network with a lot of hosts and some redundant paths, the network should be divided into several data-link network segments, which are interconnected by routers. It would be recommended that each data-link network segment should accommodate only one IP subnet (a network layer segment). This is the current network architecture in the Internet. And, both from the distributed application's point of view and from the network management's point of view, the data-link network segment (including ATM-LAN) should have a broadcast capability. This is because a broadcast is the only way to communicate with other end-stations (or routers) without the prior knowledge of their addresses.

Figure 1 shows the internet model. ATM-LAN is composed of switch nodes, and the interface with the adjacent network segment is UNI (User Network Interface), even when the adjacent network is an ATM network. Data-link network segments (including ATM-LANs) are interconnected by routers, and each data-link network segment corresponds to an IP subnet. Regarding ATM networks, since the interconnecting interface between ATM-LANs is UNI, each ATM-LAN can operate independently.

The key functionalities of a router interconnecting ATM networks are as follows.

- Cell Relaying
  Cells are relayed as ATM layer's function. Since the assignment of VPI/VCI value is performed independently at each ATM-LAN, the router has the VPI/VCI translation functionality (the procedure to establish the VPI/VCI translation table is discussed below). Since IP processing can be bypassed by ATM layer cell relaying at the router, the transmission latency to pass through the router can be sufficiently small, which is same as an ATM

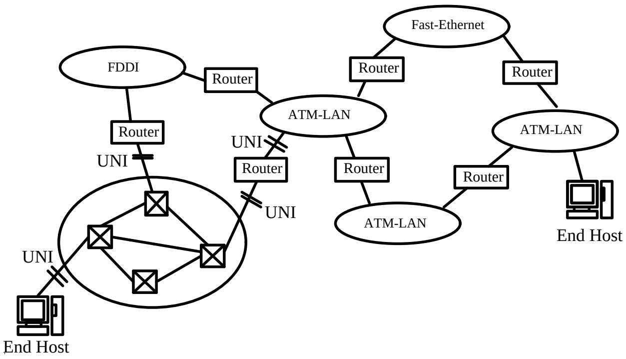

Fig.1 ATM Internet Architecture

switch node.

- End-to-end connection oriented data transfer channel set-up
  During a set-up procedure of an end-to-end connection oriented data transfer channel, e.g.RSVP with SIPP, routers co-operate with the connection (ATM-VCC as data-link connection) establishment servers in the ATM-LANs. And, also, we can achieve higher throughput than the conventional hop-by-hop IP forwarding can.

  An end-to-end connection oriented data transfer channel, that is composed of the concatenated ATM-VCCs, is established using a network layer address (i.e. IP address) rather than using an ATM address. An ATM address is used only for the establishment of ATM-VCC within a single data-link segment (i.e. ATM-LAN) through Q.2931 protocol. When a router receives the establishment request for an end-to-end connection oriented data transfer channel, it selects the appropriate next hop router (or the destination end-station) and requests the establishment of ATM-VCC toward the next hop router (or toward the destination end-station). Here, in order to resolve the ATM address of the next hop router (and of the destination end-station), the router maintains two kinds of mapping information ; (1) between the next hop router identifier and the IP address of the next hop router, and (2) between the IP address of the next hop router (and of the destination end-station) and the ATM address of the next hop router (and of the destination end-station). The former information is obtained by the routing protocol, and the later information is usually maintained at the router as a routing cache. When the later information regarding the target entity (the next hop router or the destination end-station) is missed by the routing cache, an ARP (Address Resolution Protocol) procedure [5][11] will be performed to resolve the ATM address of target entity.

  The next hop router forwards the received channel establishment request, for the end-to-end connection oriented data transfer channel, to the appropriate entity (the next hop router or the destination end-station). During a ATM-VCC establishment procedure, the ATM-LAN will indicate a VPI/VCI value for the established ATM-VCC to the router. Since the ATM-VCC establishment requests are issued to every ATM-LAN along the selected end-to-end path, each ATM-LAN will assign the different VPI/VCI code point for each ATM-VCC. Finally, a router will establish a VPI/VCI mapping table using the VPI/VCI values for these ATM-VCCs indicated by the ATM-LAN, in order to concatenate ATM-VCCs at the router.

- Security Functionality (if necessary)
  Since every cell (or could be say packet) passes through the router, the segmentation of subnet can be fulfilled as well as the current Internet architecture. For instances, the proposed router can equip security functionalities, e.g. packet filtering in firewall functions. The proposed router can also equip other security functionalities, that is, unauthorized packets are filtered just as in the case of usual firewalls and unauthorized reservation request packets for cell relaying are rejected.

As discussed above the proposed router is different from an ATM switch, because an ATM switch does not have IP forwarding capability. And, also, an ATM switch establishes an ATM-VCC based on ATM addresses with Q.2931 protocol, while a router handles IP packet flows based on IP address.

Routers are network entities which inter-connect several data-link segments. Routers exchange routing information and maintain routing information tables to forward received packets. Logically, each packet is relayed to the optimal interface by looking up the routing tables. But, actually, routers don't have to consult with the routing tables everytime they receive a packet. As an implementation, routers can have an internal cache or a bypass to minimize the packet routing delay of complex table looking-up. Likewise, routers don't have to consult with the routing table everytime they receive a cell. In an implementations, it is possible to have an internal bypass to remove the packet reconstruction delay and packet routing delay of complex table looking-up.

One possible implementation of the proposed router would be combining the ATM switch module with the conventional IP router module. And, the IP router module will be able to control the ATM switch module to establish an IP level end-to-end connection oriented data transfer channel. Having an ATM switch could be equivalent to having a cache process, that is a well-known technique applied in the current routers, from the view point of the IP router module. Finally, as discussed in sections 3 and 4, the proposed architecture does not require fundamental modifications to the routing protocol running among the routers.

Through bypassing the conventional IP forwarding process using cell-relaying, we could dramatically reduce both the IP packet processing delay and the queuing delay at the router. The IP packet processing delay is not reduced for a non-bypassed IP packet flow, but is reduced for a bypassed IP packet flow because the complex table looking-up at the router can be omitted. The queuing delay at the router for a non-bypassed IP packet flow is also reduced, since the total number of IP packets dealt with by each router (especially for core intermediate routers) will be reduced. In the current Internet, the observed IP packet delay at the routers was about from few *msec* to few *sec*, when we fulfill a traceroute. On the contrary, the expected cell delay at ATM cell forwarding module could be from few μsec to *msec*.

3. Connection Oriented Service Architecture

Currently, the connection oriented IP transport flow (i.e. a network layer connection oriented service) has been discussed in IETF, e.g. [7][10]. The goal of connection oriented communications is to provide an end-to-end IP transport. Such a transport flow can have a certain QOS. In such IP transport flow, there is an admission policy at the network layer level, that is a similar concept to CAC (Connection Admission Control) in ATM networks. When the IP transport flow set up request is admitted, the resource for the transport flow is reserved at corresponding routers. Here, in order to find out the next network (or can be said the router) that the IP transport flow should be routed to, some kind of routing protocol (e.g. similar to OSPF) is executed among the routers.

Unlike other models of IP over ATM, the architecture proposed in this section assumes QOS-ed communications over not only ATM but also the other types of platforms, using the concept of connection oriented IP transport. The connection oriented IP transport is mapped into data-link layer connection (ATM-VCC) within the ATM networks.

Connection oriented data-link networks (e.g.ATM) can provide a QOS-ed connection between all the data-link SAPs (Service Access Points) within the same data-link network segment through a connection set up procedure . A communication beyond data-link network segment is performed by a QOS-ed IP transport flow to the router. When the data-link network segment is connection oriented, a router or a source end-station establishes a QOS-ed data-link connection (e.g. ATM-VCC) toward the next router. The next router performs the same procedure, until the IP transport flow setting up message reaches at the destination end-station.

After the IP transport flow is established, IP packets are forwarded through the QOS-ed pipe. A unique identifier, called flow-ID (flow identifier), could be used along the IP transport flow [7][10]. Within the ATM network, VPI/VCI can be used instead of flow-ID defined in the IP layer.

The IP packet forwarding procedure is as follows.

1. A packet arrives at router.
2. The flow-ID and the IP address (source / destination) are checked and the next hop router is determined.
3. TTL (Time To Live) is decremented.
4. Unless this router is the destination end-station, the packet is forwarded.

This procedure is not ATM-specific and it is applicable to all other platforms. Three cases are discussed below.

- ATM-LAN → ATM-LAN
  The router has a mapping table between the ingress VPI/CPI and the egress VPI/VCI for each cell flow according to the appropriate IP transport flow. Then, it is unnecessary to examine the flow-ID and IP address in step (2) : "Re-direction" of the ingress and the egress data-link connection. Here, the mapping table is established during the connection set up procedure.
  1. A cell arrives at a router
  2. VPI/VCI of the ingress cell is examined and the next hop VPI/VCI of the egress cell is determined (maybe by hardware)
  3. TTL in the IP header may not be decremented
  4. Unless this router is the destination end-station, the cell is forwarded.
- ATM-LAN → Other-LAN
  A router will just reassemble a IP packet to forward it to the appropriate router or end-station.
  1. A cell arrives at a router
  2. An IP packet is reassembled (e.g. AAL5)
  3. The flow-ID and the IP address (source / destination) is examined and the next hop router is determined.
  4. TTL in the IP header is decremented

5. Unless this router is the destination end-station, the packet is forwarded.

- Other-LAN → ATM-LAN
  A router has a mapping table between the ingress flow-ID/IP-address and the egress VPI/VCI for each IP transport flow. Here, again, the mapping table is established during the connection set up procedure. Here, the TTL decrement value in step 3 may not be one. This is because, regarding the IP packets belonging to the cell-relaying bypass routes, the TTL in the IP packet header is not dealt with at the intermediate router(s).

  1. A packet arrives at a router
  2. The flow-ID and the IP address (source / destination) is examined and the next hop VPI/VCI of the egress cells is determined.
  3. TTL in the IP header is decremented (may not be one)

When ATM end-stations communicate purely over ATM-LANs, a seamless end-to-end cell-relaying virtual channel can be established between them. It should be noted that, even though many ATM-VCCs (data-link connection) are used at the data-link layer, the seamless connection over ATM-LANs is at a network layer connection (not a data-link layer connection).

4. Connectionless Service Architecture

4.1 Connectionless Services over ATM Networks

A connection oriented service must perform an IP level connection set up procedure, before transferring the information. It is said that a connection set up latency for a large scaled ATM network would not be small [4]. On the contrary, a connectionless service does not require a connection set up procedure, basically. ATM-VCC (a data-link connection), in order to provide a connectionless service, has been established in advance. Here, ATM-VCCs to provide a connectionless (conventional IP forwarding) service would be SVC (Switched VC) rather than PVC (Permanent VC). Once the VC(s) for the connectionless service is(are) established, there will be no connection set up latency to transfer connectionless IP packets. Therefore, even when ATM networks and other networks provide an end-to-end connection oriented data transmission service, a connectionless service will be still important. And, from the view point of co-existence with the network that is not connection oriented (e.g. FFOL), a connectionless data transmission service (i.e. a conventional IP forwarding) must be provided by ATM networks.

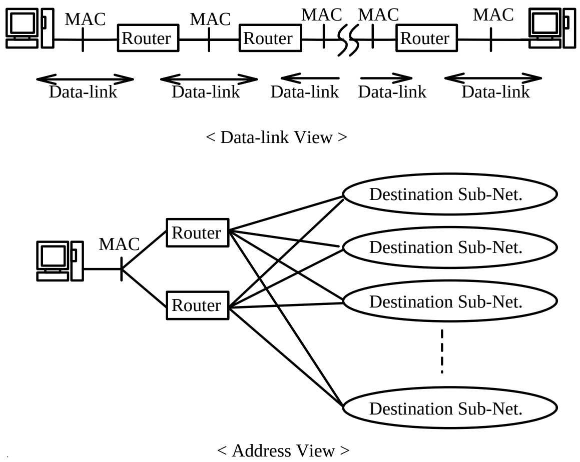

Fig.2 Conventional IP Networks

4.2 Connectionless Services in Conventional Data Networks

In IP networks, IP processing will be generally performed, whenever the IP packet is transferred to the other data-link network segment [12]. This means that, in a large scaled network, the IP packet will experience many IP processing points between source and destination end-stations. For example, when the IP packet is transferred from Toshiba's gateway host (*"isl.rdc.toshiba.co.jp"*) to a certain host in France (*"chorus.chorus.fr"*), an IP packet experienced 19 hops (IP processing) to reach the destination host. Therefore, in the large scaled network, the large delay for IP packet delivery is expected, due to a large number of IP processing points. This is similar to the CLS (ConnectionLess Server) approach defined by ITU-T [2].

Figure 2 shows data-link view and address view in an IP based network. Here, each router can recognize the destination subnet (i.e.Net-ID) of the received IP packet, as shown in address view. However, the network layer protocol (e.g. IP) is performed at every router to terminate data-link layer (i.e. MAC). As a result, even when the distance (i.e. the number of IP subnets to reach the destination IP subnet) associated with the address management view is small, the distance associated with the data-link layer's view is not small.

4.3 Proposed Connectionless Service Architecture

There is no difference between the connectionless IP packet transmission in the proposed architecture and that of existing IP forwarding. Within the data-link network segment, end-stations and routers exchange IP packets using data-link connections (e.g.ATM-VCCs). IP packets beyond the data-link segment are sent to an appropriate router from the source end-station. If a router has some knowledge on a local topology (e.g. the topological information in the routing domain), it is possible to bypass IP layer processing at the intermediate routers, by having the concatenated ATM connections between distant routers.

Routers exchange routing information and forward IP packets to the appropriate router. The following discussion is regarding the network that has only ATM-LANs. Regarding toward the other types of platform from an ATM network, ATM-VCC shall be always terminated at the router to execute the conventional IP forwarding process. On the contrary, regarding toward ATM network from the other types of platform, you can see the router as the end-station in the following discussion.

The proposed ATM router has both cell relaying and IP forwarding capabilities. An IP packet toward out-side the ATM-LAN where the source end-station belongs to is sent to an appropriate router from the source end-station. Routers exchange routing information and forward packets to the appropriate router. Regarding the current routing protocol (e.g. OSPF), every router does know all of subnet IDs (Net-IDs) within the domain where the routing protocol is applied.

4.3.1 Hop-by-hop ATM-VCC cacheing for active IP flow

When the first IP packet (of the session or after the idle period) arrives at a router (attaching to ATM), the next router is decided using the IP header, i.e. the routing decision will be done based on the IP address (and the flow-ID in IPv6). After the next router is decided, a dedicated ATM-VCC is picked up for the IP packet flow. When the packet comes from an ATM network (or a connection-oriented platform, e.g. Frame Relay), the router maps between the incoming VPI/VCI (or some flow-ID in the data-link layer)and the (picked up) outgoing VPI/VCI. When the packet comes from the conventional platform (e.g. Ethernet/FDDI), the router always examines the IP header in the case of IPv4 (in IPv6, the flow-ID could be mapped to the outgoing VPI/VCI). When the IP packet comes from an ATM network, the successive IP packet is forwarded based on the incoming VPI/VCI value, without any usual IP processing.

If the activity of the IP packet flow is/becomes low, the mapping information (i.e. the cached information) is pushed out. This means that the pushed out IP packet flow from the cache must be examined by the usual IP processing (looking at the IP header) at the router. As a result, IP packet flow with a high activity can bypass a usual IP processing at the router.

The above procedure is performed at every router (e.g. every intermediate router that interconnects ATM networks). Then, an intermediate router can deal with a high activity IP packet flow with a very small latency. This is because IP packet forwarding is done by cell relaying, which will be done by a hardware. The job of router is maintaining the VPI/VCI mapping table in the cell switch, or is maintaining the mapping table between the flow-ID and the VPI/VCI, as a cache (bypass) processing for the usual IP forwarding process.

When every data-link segment is an ATM network, small latent IP packet forwarding will be provided, even for a connectionless IP packet. This is because an IP packet with a high activity is relayed on the cell-by-cell bases at every router.

4.3.2 Bypassed ATM VCC for active transaction

When a router knows that it will have or is having a lot of communication with a certain router (or a subnet), it may create a cell-relaying bypass route which seamlessly connects the two routers at cell level. This cell-relaying bypass route cuts through the IP processing at the intermediate routers along the packet transmission path between the two routers. On the other hand, when a router does not use the cell-relaying bypass route for IP packet forwarding, the IP packet will be forwarded hop-by-hop, that is completely the same approach as the conventional IP forwarding.

When a router can establish a seamless cell-relaying route bypassing the intermediate router's network layer protocol processing along the routing path to the destination end-station, we can reduce the end-to-end IP packet delivery latency due to reducing the number of network layer protocol processing points at the intermediate routers. Here, it is unnecessary to propagate the information of such cell-relaying bypass routes for every router. The information of cell-relaying bypass routes shall be maintained by the source and destination routers associated with the cell-relaying bypass routes. This means that the routing protocol and the routing information exchanged by routers is nothing different from the existing routing protocol.

The information and the path to bypass the intermediate router(s) could be seen as a cache image associated with the information and path for hop-by-hop routing (i.e.conventional IP forwarding routing). When the cache is missed (i.e. when there is no bypass route), the hop-by-hop information is referenced and the IP packet is forwarded to the adjacent router. On the contrary, when the cache is hit (i.e. when a bypass route exists associated with the destination subnet), the cache information is referenced and the IP packet is forwarded using the bypass route.

The establishment of the cell-relaying bypass route will be issued by the following two cases. Here, the establishment procedure of the cell-relaying bypass route is the same as the establishment procedure of the end-to-end connection oriented data transfer channel set-up discussed in section 2.2. And, the source router and the destination router discussed in this subsection can be an intermediate router, from the view point of end-stations. This means that the source (or the destination) router may not directly connect to the source (or destination) end-station through a single cell-relaying bypass route. In other words, in order to reach the source router from the source end-station (or to reach the destination end-station from the destination router), intermediate router(s) may be required.

1. Source router
   The source router has the activity information regarding the destination subnet's. When the IP forwarding activity toward the certain destination subnet is high, the source router establishes the cell-relaying bypass route toward the destination subnet's router. When the activity becomes low, the established cell-relaying bypass route will be torn down (i.e. the cache information is pushed out).

   Every router would establish the cell-relaying bypass route based on the IP flow's activity information. When there is no merged IP packet flow along the packet transmission path associated with the high activity IP packet flow and all of the IP packets belonging to the high activity IP packet flow take the same path from the source router to the destination router, every router along the packet transmission path will try to establish the cell-relaying bypass route to the same destination router. However, in general, we could say that there are merged IP packet flows, that have the same destination router, along the high activity IP packet flow. And, we could say that, in hop-by-hop IP forwarding, the IP packet forwarding path would not be the same for all the IP packets. Also, even when some routers (source router and intermediate routers) establish cell-relaying bypass routes to the same destination router simultaneously, only one bypass route will finally survive. The survived bypass route will be the route that is established by the farthest router from the destination router. The bypass routes from the other routers will be automatically torn down (i.e. cache information will be pushed out), since the activities of the IP packet flow toward the destination router will be reduced.

   A cell-relaying bypass route from the source router toward the destination router can be established without the activity information. This means that bypass route can be established both dynamically and statistically based on the network's topology information. And, of course, whether to establish the bypassed route based on the activity information depends on each router's policy.

2. Intermediate router
   When the intermediate router becomes high loaded, the intermediate router can create bypass routes that is bypassing itself. The intermediate router establishes the cell-relaying bypass route for the active source and destination router pair. After the establishment of two cell-relaying routes, that are (1) from the source router to the intermediate router and (2) from the intermediate router to the destination router, these two cell-relaying routes are coupled at the intermediate router. This procedure could be seen as a re-direction of cell-relaying : i.e. the cell-relaying coming from the source router to the IP forwarding entity in the intermediate router is "re-directed" to the cell-relaying entity toward the destination router.

   The cell-relaying bypass routes will be torn down by the intermediate router, when the processing load at the intermediate router becomes low. Generally, the threshold value of the processing load to tear down the cell-relaying bypass route(s) should be smaller than the threshold value to create the the cell-relaying bypass route(s), in order to avoid the oscillation tendency. As a result, the cell-relaying bypass route(s) established by the high loaded intermediate router would be always between a high active source and destination router pair, when the intermediate router becomes high loaded.

The ideal and optimal case is where full meshed cell-relaying bypass routes are established among all the routers that exist in the routing entry of the routing information. In this case, the IP packet delivery will be performed as shown in figure 3. In the following subsection, the evaluation regarding the required address points (VPI/VCI resource) to provide a connectionless service at the ideal/optimal case is performed. The evaluation below gives us the worst case regarding required address points to provide a connectionless service, and also gives us the best case regarding the maximum number of actual network layer protocol processing points.

Obviously, the every host can act as the same as the router discussed above (i.e. you can see the router as a end-station). Then, the seamless cell transmission channel will be established, i.e. (a) from the source end-station to the router, (b) from the source end-station to the destination end-station, and (c) from the router to the destination end-station.

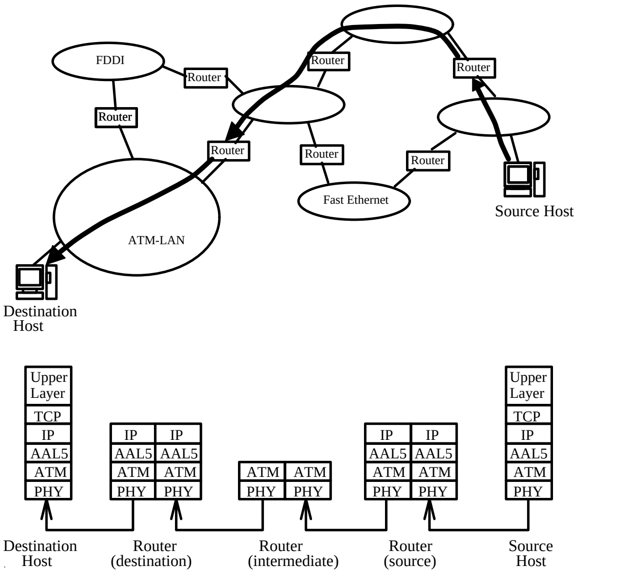

Fig.3 Proposed IP Packet (Connectionless) Transfer

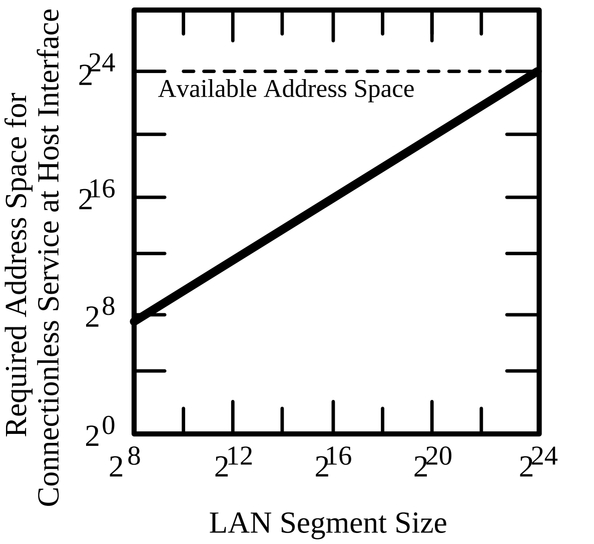

Fig.4 Required Address Space at End-station's UNI

4.4 Performance Evaluation

In this section, (1) the required VPI/VCI code points (VPI/VCI resource) to provide connectionless service and (2) the maximum number of actual network layer processing points for a large scaled network are evaluated. The evaluation is performed assuming the ideal and optimal case, where full meshed cell-relaying bypass routes are provided among all the routers. Here, since the ATM header of the UNI cell format has 24 bit of VPI/VCI field, the theoretically available VPI/VCI code points (VPI/VCI resource) are 2^{24}.

4.4.1 Single Layered Network

Figure 4 shows the required address points for a connectionless service at an end-station versus the LAN segment size. Here, the LAN segment size means the number of accommodated end-stations in the ATM-LAN. And, regarding an intra-ATM-LAN IP packet delivery, a Full-Mesh method is assumed [13][14]. When the LAN segment size is not too large, say less than 2^{12}, the required address points (VPI/VCI resource) for a connectionless service can be sufficiently small. When the LAN segment size is 2^{12}, the required address points is less than $0.025\%(1/2^{12})$ of the available address points. This means that more than 99.9% of address points (VPI/VCI resource) has still remained to be used for a connection oriented service. Here, 2^{12} is four times larger than the capacity of Ethernet (2^{10}).

Figure 5 shows the required address points (VPI/VCI resource) for a connectionless service at a router regarding incoming ATM-VCCs from the other routers and end-stations to the router versus the number of accommodated end-stations in the ATM internet. "m" represents the LAN segment size, which is the number of accommodated end-stations within the ATM-LAN segment. The required address points is approximately given by $m + N/m$. Here, "N" represents the number of accommodated end-stations in the ATM internet. Figure 6 shows the required address points (VPI/VCI resource) for a connectionless service at a router regarding outgoing ATM-VCCs from the router to the other routers and the end-stations versus the LAN segment size. The number of required address points for a connectionless service corresponds to the number of the network addresses (e.g. Net-IDs) maintained by the router, which is given by N/m. Here, we assumed that every LAN segment size is identical.

Theoretically, the proposed architecture can accommodate a huge number of end-stations, that is about up to $2^{48}(=2^{24} \times 2^{24})$, with only two network layer protocol processing points for a connectionless service. This roughly corresponds to the IGP/BGP separation of the current Internet. When the LAN segment size is larger and/or the number of LAN segments in the ATM internet is larger, the ATM internet can accommodate a larger number of end-stations. However, we may not be able to implement such a large scaled ATM internet by a single layered structure. One is due to the capability regarding the number of maintained addresses in the router. And the other is due to the LAN segment size. Yet another is due to the load concentration of the global traffic to the connectionless servers. When the number of LAN segments is 2^{24}, the number of maintained addresses regarding a connectionless service in a router for an Intra-LAN IP packet delivery

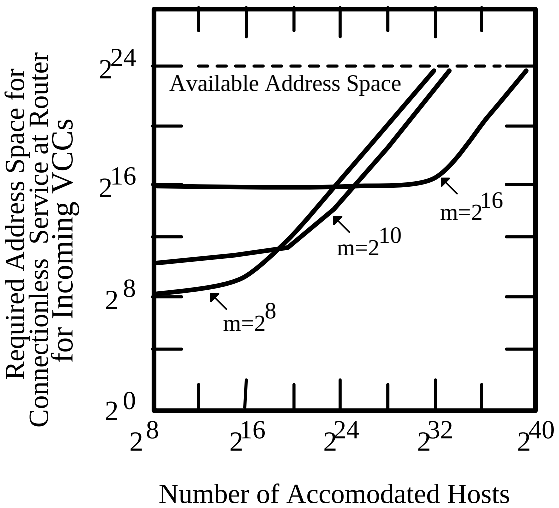

Fig.5 Required Address Space at Router

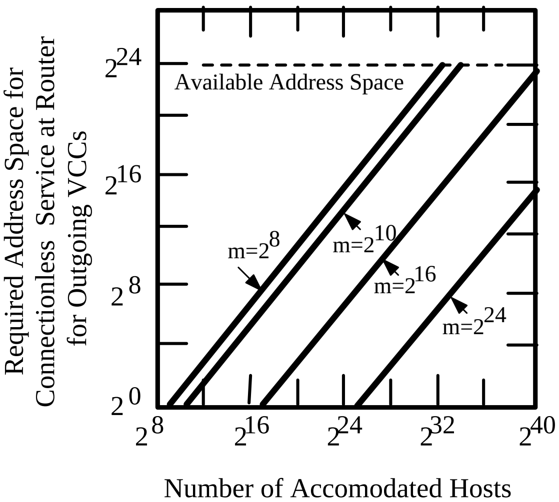

Fig.6 Required Address Space at Router

is 2^{24}. That seems to be too large to maintain. Here, when the number of LAN segments is 2^{24}, the number of routing entries maintained in the router corresponds to the case where all the subnets defined in the current Internet are class C. 2^{24} of LAN segment size seems to be too large to be accommodated by a single ATM-LAN. Therefore, below, the feasible size of the ATM internet to be implemented is assumed.

It is assumed that the number of accommodated end-stations in the ATM internet is about up to 2^{23}. When the LAN segment size is 2^{10} (the same as the Ethernet capacity), the required address points for a connectionless service at a router is about $2^{13}(\simeq 2^{23}/2^{10} + 2^{10})$. Here, 2^{13} of address points are less than 0.05% ($\simeq 1/2^{11}$) of the available address points. This means that more than 99.95% of address points (VPI/VCI resource) still remain to be applied for a connection oriented service.

4.4.2 Scaling Up Method and Evaluation

IPng (IP next generation) has a 128 bits address space [8], and E.164 has an up to 15 digits ($\simeq 2^{50}$) address space. Here, as mentioned in [9], the target number of accommodated end-stations in the Internet based on IPng (IPv6) is 10^{15}, which is the same as the maximum number of end-stations for the network based on E.164 addressing. In order to apply to such a large scaled network, the hierarchical internetworking of the proposed architecture is discussed and evaluated. Figure 7 shows the two layered structure of ATM internet. The SAP (Service Access Point) of the higher layer internet corresponds to the lower layer internet. And the SAP of the lower layer internet corresponds to the end-station. This means that, when the number of accommodated SAPs in the ATM internet of each layer is N, the n layered ATM internet can totally accommodate N^n end-stations (SAPs). The maximum number of network layer protocol processing points is given by $2 \times (2n - 1)$.

When the structure of hierarchical ATM internet corresponds to the (hierarchical) addressing structure, the routing of the connectionless IP packet would be easier. Both addressing structure of ITU-T (E.163 and E.164) and of ATM Forum (based on GOSIP address) are hierarchical and well-managed for easy routing of the IP packet. Also, IPv6 (i.e. SIPP) can be hierarchical.

Below, two example for well-known large networks are discussed and evaluated.

- IP version 4 (IPv4)
  Two layered ATM internet will be feasible to cover. IPv4 has a 32 bits address space and the size of the defined network segment is generally 2^{24}, 2^{16} or 2^8, excluding CIDR addressing. In order to align to

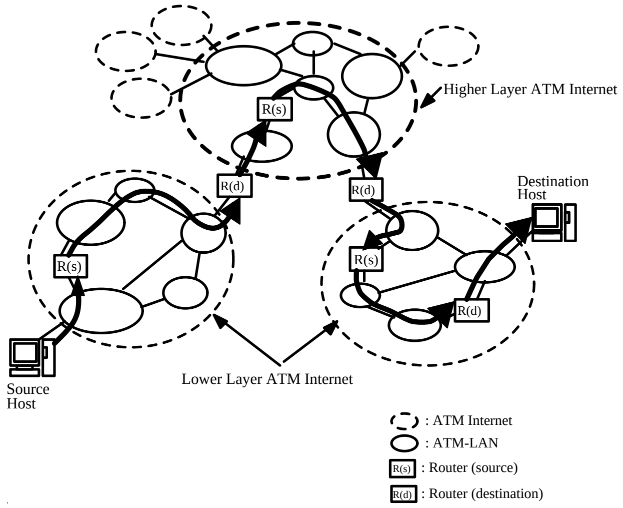

Fig.7 Hierarchical Scaling Up Approach

the structure of the network segment size in IPv4, the LAN segment size would be $2^8 \times n$. Here, $2^8 \times n$ should be feasible value to be implemented as an actual ATM-LAN segment size. When both the higher layer and lower layer accommodate 2^{16}, both the LAN segment size and the number of LAN segments are 2^8. Here, the LAN segment size is defined as the number of accommodated SAPs in the LAN segment. The required address points for a connectionless service at a router is about $2^9 (\simeq 2 \times 2^8)$, that is $1/2^{15}$ ($\simeq 0.003\%$) of the available address points as UNI. This means that more than 99.99% of address points (VPI/VCI resource) are available for a connection oriented service. And, the number of maintained addresses regarding a connectionless service in the router is small enough, which is only about 2^8.

Then, the maximum number of actual network layer protocol processing points regarding an end-to-end connectionless IP packet delivery is six.

- E.164 (numbering plan for ISDN)
  E.164 has a 15 digits ($\simeq 2^{50}$) address space. The country code will be 2 digits which is the same as the E.163 addressing. The coding structure of following 13 digits ($\leq 2^{44}$) may not be defined, yet. Since, we could accommodate up to around 2^{44} ($=2^{22} \times 2^{22}$) end-stations by the two layered network, the three layered network would be able to apply to E.164 network. One example is as follows. The lowest layer has a seven digits ($\simeq 2^{24}$) address space, which is the same as the current telephone numbering in USA. The second layer has a six ($\simeq 2^{20}$) digit address space, and the highest layer has two digits (a country code as it is).

  Since the lowest layer has the largest number of accommodated SAPs, the lowest layer case is evaluated. 7 digits would be divided into 3 digits and 4 digits, as well as the current telephone numbering. Then, we can cover this structure using the following network segment size. The LAN segment size is $2^{14} (\simeq 10^4$; 4 digits), and the number of LAN segments is $2^{10} (\simeq 10^3$; 3 digits). The required address points for a connectionless service at router are about $2^{14} (\simeq 2^{14} + 2^{10})$, which is about 0.1% ($1/2^{10} = 2^{14}/2^{24}$) of the available address space. This means that more than 99% of address space (VPI/VCI resource) is available for a connection oriented service. And, the number of maintained addresses regarding a connectionless service in the router is small enough, which is only about 2^{10}.

  Then, the maximum number of actual network layer protocol processing points regarding an end-to-end connectionless IP packet delivery is ten.

  Here, since the target number of accommodated end-stations in the network based on IPv6 is also 10^{15} [9], three layered ATM network could also cover an IPv6 network. But, the arrangement of network size (the number of SAPs and the number of subnets) for each layer will be different from E.164 addressing structure's case.

5. Conclusion

A high throughput small latent IP packet delivery architecture using ATM technology is proposed in this paper. Using the ATM's cell relaying capability at a router, we can provide a high throughput small latent end-to-end IP packet delivery.

Regarding connection oriented IP packet transmission, since the router can map between the flow-ID (and IP address) in the IP packet header and the VPI/VCI, we can reduce the IP processing latency at the routes within the ATM-LANs.

Regarding connectionless IP packet transmission, we can also reduce the IP processing latency at the routers within the ATM-LANs, through the re-direction (bypassing) technique at the routers. When the proposed architecture is applied, we can reduce the processing load of each router, as well as reducing the end-to-end packet delivery delay. This is because, due to bypassing the IP processing by re-direction technique, the amount of IP processing for intermediate IP packets can be reduced. By the proposed architecture, even when the number of accommodated end-stations in the network is 10^{15} (the E.164 capacity and the target of IPv6), the maximum number of actual IP processing points for a connectionless service can be only ten. And, more than 99% of address space (VPI/VCI resource) can remain for a connection oriented service.

Acknowledgement

The authors thank to many ATM experts in JAIN consortium in Japan for many discussions regarding the proposed architecture in this paper. And, the authors appreciate many discussions with Mr.Oumar Ndiaye, who belonged to MIT in USA, regarding the software implementation of the proposed architecture.

References

[1] ITU-T Rec.I.150, "BISDN Asynchronous Transfer Mode", 1990.

[2] ITU-T Rec.I.364, "Support of Broadband Connectionless Data Service on BISDN", June, 1992.

[3] ATM Forum, "ATM UNI Specification 3.0", Aug., 1993.

[4] A.Schmidt, R.Campbell, "Internet Protocol Traffic Analysis with Applications for ATM Switch Design", *Computer Communications Review*, vol.23, No.2, pp.39-52, April, 1993.

[5] M.Laubach, "Classical IP and ARP over ATM", *IETF RFC*, RFC1577, January, 1994.

[6] J.Heinanen, "NBMA Address Resolution Protocol", *IETF Internet-Draft*, draft-ietf-atm-nbma-01.txt, June, 1993.

[7] S.Deering, "Simple Internet Protocol Plus (SIPP) Specification (128-bit address version)", *IETF Internet-Draft*, draft-ietf-sipp-spec-01.txt, July, 1994.

[8] P.Francis, S.Deering, R.Hinden, R.Govindan, "Simple Internet Protocol Plus (SIPP) : Addressing Architecture", *IETF Internet-Draft*, draft-ietf-sipp-routing-addr-02.txt, July 1994.

[9] F.Kastenholz, C.Partridge, "Technical Criteria for Choosing IP:The Next Generation (IPng)", it IETF Internet-Draft, draft-kastenholz-ipng-criteria-02.txt, May, 1994.

[10] L.Zhang, R.Braden, "Resource ReSerVation Protocol (RSVP)", *IETF Internet-Draft*, draft-ietf-rsvp-spec-02.ps, May, 1994.

[11] M.Ohta, H.Esaki, K.Nagami, "Conventional IP over ATM", *IETF Internet-Draft*, draft-ohta-ip-over-atm-01.txt, July, 1994.

[12] D.E.Comer, *"Internetworking with TCP/IP"*, Prentice Hall, 1991.

[13] H.Esaki, Y.Tsuda, T.Saito, S.Natsubori, "Class D Service Architecture in ATM-Internet", *IEEE ICC'94*, May, 1994.

[14] H.Esaki, Y.Tsuda, T.Saito, S.Natsubori, "Datagram Delivery in an ATM-Internet", *IEICE Trans. on Communications*, Special Issues for Future Private Networks, Vol.78-B, no.3, March, 1994.

Hiroshi Esaki received the B.S. and M.S. degrees from Kyushu University, Fukuoka, Japan, in 1985 and 1987, respectively. In 1987, he joined Research and Development Center, Toshiba Corporeation, where he engaged in the research of ATM systems. He has been at Bellcore in New Jersey (USA) as a residential researcher from 1990 to 1991, and has engaged in the research on high speed computer communications. From 1994, he has been at CTR (Center for Telecommunications Reserch) of Columbia University in New York (USA) as a visiting scholar. He is currently interested in an internet architeture using ATM technology and a mobile computing.

Masataka Ohta recieved the B.S. and M.S. degrees from University of Tokyo, Japan, in 1982 and 1984, respectively and the Ph.D. (Science) degree from Tokyo Institute of Technology, Tokyo, Japan, in 1994. He has been working at Computer Center of Tokyo Institute of Technology since 1987. He is a member of IPSJ, ACM and ISOC. He is currently interested in Internet technology including multi-lingual communication and ATM technology, high performance computing and computer graphics.

Ken-ichi Nagami received the B.S. degree from Chiba University, Japan, in 1990 and M.S. degree from Tokyo Institute of Technology, Japan, in 1992. He joined Research and Develelopment Center, Toshiba Corporation where he has been working on communication system. He is currently interested in internet technology for high speed networks.