PAPER

Flow Attribute Notification Protocol (FANP) for Label Switching

Ken-ichi Nagami† **Yasuhiro Katsube** † **Yosuro Shobatake** † **Akiyoshi Mogi** † **Shigeo Matsuzawa** † **Tatsuya Jinmei** † *and* **Hiroshi Esaki** † *Members*

SUMMARY This paper proposes and performs the primary feasibility evaluation on Flow Attribute Notification Protocol (FANP), which is a protocol between neighbor CSR (Cell Switch Router) nodes for the management of cut-through packet forwarding, in order to apply label switching paradigm. In cut-through packet forwarding with label switching, a router doesn't have to perform conventional IP packet processing for the received packets. FANP indicates the mapping information between a data-link connection and a packet flow to the neighbor node. FANP defines two key procedures, i.e., one is the VCID Notification Procedure, and the other is the Flow-ID Notification Procedure. The VCID Notification Procedure lets the label switching paradigm over the label swapped data-link, such as ATM link, though the other label switch architecture can not work over the label swapped data-link.

The primary evaluation of FANP has been performed using the prototype system and with the actual packet statistics. The result shows that, with a corporate backbone level, the label switch router system with FANP would work well.
key words: Label Switching, FANP, ATM, CSR, cut-thru, TCP/IP

1. Introduction

Due to many reasons, e.g., scalability requirement and exclusive resource allocation to provide QoS (Quality of Service) and CoS (Class of Service), the connection oriented (CO) data-link platforms, e.g., ATM and Frame Relay, are going to be used as well as legacy connection less (CL) data-link platforms, e.g., Ethernet and FDDI. One of the important features of the CO data-link is the presence of a data-link-level connection identifier. In the CO data-link, we can establish multiple virtual connections (VCs) with their VC identifiers among the nodes. When we aggregate packets that have the same direction (e.g., having the same destination IP address) into a single VC, we can forward the packets in the VC without IP processing. With this configuration, routers can decide which node is the next-hop for the packets based on the VC identifier. In the LSR (Label Switch Router) architecture, a label (e.g., VPI/VCI for the ATM interface), is used for packet forwarding label according to given packet flow. Using the label for packet forwarding information, router does not need to analyze the full IP address in the received IP packet [1][2][3][4]. Therefore, we can achieve higher and more flexible packet forwarding capability than the conventional routers can. In order to apply the label switching paradigm, we have to apply a new protocol, e.g., IFMP [5], TDP[4]. With a new protocol, LSR can forward the incoming packets using an ATM switch engine bypassing the conventional IP processing. According to the ingress VPI/VCI value with ingress interface information, LSR determines the egress interface and the egress VPI/VCI value. CSR (Cell Switch Router) is one of the label switch router architectures, discussed at the MPLS-WG (Multi-Protocol Label Switch Working Group) in the IETF [3][8][9][10]. CSR uses the FANP (Flow Attribute Notification Protocol) [6], discussed in this paper, in order to apply the label switching paradigm. With FANP, the label switching paradigm can apply to the ATM link, though the other protocols for each label switch architectures can not.

The purposes of this paper are (i) clarifying the key (and unique) function of FANP, comparing to the protocols for the other label switching architectures, and (ii) performing the primary feasibility evaluation. Section 2 describes the requirement and the uniqueness of FANP. Section 3 gives the brief protocol overview of FANP. Section 4 gives the primary feasibility evaluation of FANP implementation. Section 5 gives a brief conclusion.

2. Requirements and Design Principles

2.1 Issue of LSR Protocols

Protocols, including FANP, for LSR (Label Switch Router) are proposed [1] [5] [6] [7]. The unique feature of FANP for CSR architecture, comparing with other protocols, is applying the label switching paradigm over the label swapping data-links, such as the ATM links. In the label swapping data-links, the ingress and egress data-link label (i.e., VPI/VCI for ATM link) is generally the different value.

ATM is one of major data-link technologies, especially in the wide area networks. When we construct the IP networks, they include ATM links, as well as

Manuscript received February 4, 1998.
Manuscript revised April 11, 1998.
†The authors are with R&D Center, TOSHIBA CORPORATION, Kawasaki-shi, 210 Japan.

the legacy digital link (e.g., SDH/PDH/SONET). In other words, the public IP networks (i.e., Internet) and the private IP networks (i.e., intranet and extranet) use ATM links to establish their IP networks. IP network technology, including label switching technology, should be able to apply to any data-link.

Protocols, excepting FANP, proposed for each particular label switch router architecture [1] [2] [4], can not apply the label switching paradigm over the ATM link. This is because the VPI/VCI value is directly mapped with the IP packet flow (i.e., Flow -ID) in these LSR architectures. The LSR (LSR1) indicates the Flow-ID and the corresponding VPI/VCI value to the neighbor LSR node (LSR2), in order to establish the mapping between the Flow-ID and the data-link label (i.e., VPI/VCI value). However, since the VPI/VCI values at the neighbor LSR node (LSR2) connected by the ATM link are generally the different value, the LSR2 establishes a wrong mapping state between the Flow-ID and the corresponding VPI/VCI value. This mechanism works, only when the ingress and egress VPI/VCI values of the ATM link are the same value. In other words, this mechanism does not work over the general ATM link nor over the label swapped data-link.

The FANP defines the VCID Notification Procedure and the Flow-ID Notification Procedure, discussed in section 3.2 and in [6], in order to apply the label switching paradigm over the label swapped data-link, such as the ATM link.

2.2 Requirements and Design Principles for FANP

The followings are the protocol requirements and design principles for FANP.

1. Applicable to various types of data-link , including label swapped data-link
2. Available with various connection types (e.g., SVC, PVC, VP)
3. Robust operation

The system should operate correctly even under the following conditions.

1. VC failure
   Some systems can detect VC failure as the function of data-link (e.g., OAM function in the ATM). However, we can not assume all nodes in the system can detect VC failure. The system has to operate correctly, assuming that every node can not detect VC failure.
2. Message loss
   Control messages in the FANP may be lost. The system has to operate correctly, even when some control messages are lost.
3. Node failure
   A node may be down without any explicit notification to its neighbors. The system has to operate correctly, even with node failure.

Though FANP is not the protocol only for ATM link but for any data-link, the following discussion in this paper assumes that the data-link is an ATM network. And, the scope of the protocol discussed in this paper is only for unicast.

3. FANP: Flow Attribute Notification Protocol

3.1 Protocol Overview

Figure 1 shows an operational overview of FANP. In the figure, a cut-through packet forwarding path is established from host 1 (H1) to host 2 (H2) using two Dedicated-VCs. H1 and H2 are connected to Ethernets, and R1, R2 and R3 are routers which can speak FANP. R1 and R3 have both an ATM interface and an Ethernet interface. R2 has two ATM interfaces. When R1 receives an IP packet from H1, R1 analyzes the payload of the received IP packet whether it is a trigger packet or not. When the received packet is a trigger packet, R1 fetches a Dedicated-VC to its downstream neighbor(R2) and sends FANP messages. FANP is effective only between the neighboring nodes. The same procedure would be performed between R2 and R3 independently from the procedure between R1 and R2. The flow-ID of the packet flow from H1 to H2 is represented as id(H1,H2). Here, id(H1,H2) is the set of the IP address of H1 and that of H2.

The Dedicated-VC is released when no packet is transferred on it for a given period. We do not need to explicitly indicate release of the Dedicated-VC to the neighbor node, since the state management in FANP is of soft-state, rather than of hard-state.

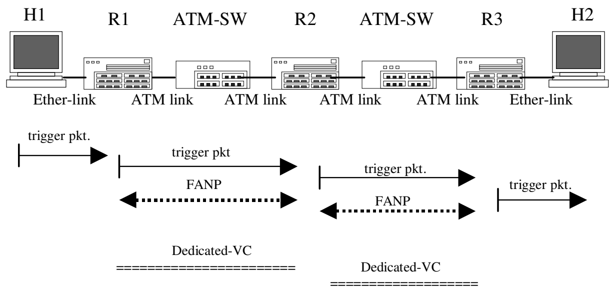

Figure 1. Trigger Packet and FANP Initiation

3.2 FANP Protocol Specification Overview

FANP has the following five procedures, that are (1) Dedicated-VC selection, (2) VCID notification, (3) flow-ID notification, (4) Dedicated-VC refresh and (5) Dedicated-VC release. Procedures (2), (3) and (4) have nothing to do with the kind of the Dedicated-VC; i.e., SVC, PVC or VP. On the contrary, the procedures (1) and (5) with SVC are different from the procedures with PVC and with VP. The detailed procedures are described in the following subsections. Figure 2 gives the message exchange for establishing and maintaining the cut-through paths with FANP system.

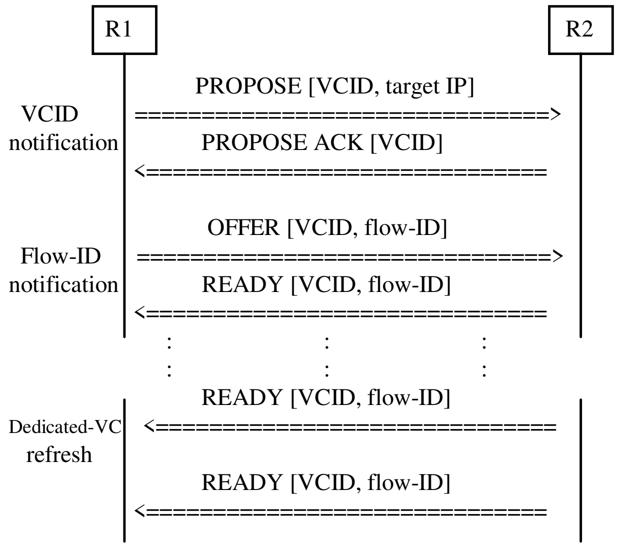

Figure 2. Message Exchange for Establishing and Maintaining the Cut-through Path

3.2.1 Dedicated-VC Selection Procedure

A VC is picked up in order to use as a Dedicated-VC. The ways of picking up the Dedicated-VC is either of the followings.

1. A number of VCs are prepared in advance, and registered into an un-used VC list. When a Dedicated-VC is needed, one of them is picked up from the unused VC list.

2. A new VC is established through ATM signaling on demand.

With ATM PVC/VP configuration, a Dedicated-VC is activated by the procedure (1). With ATM SVC configuration, a Dedicated-VC is activated by the procedure (1) or (2). When the procedure (1) is used, some number of VCs are prepared in advance through ATM signaling. These VCs are registered into the un-used VC list. When a Dedicated-VC is needed, a VC is picked up from the un-used VC list. When the procedure (2) is applied, a Dedicated-VC is always established using the ATM signaling.

The procedure (1) can decrease a time to activate a Dedicated-VC. But the necessary VC resource will increase to provide an additional VCs. Which procedure should be applied to is a matter of local decision for each physical link, taking the economical requirement and the system responsiveness into account. A Dedicated-VC is a uni-directional VC. This means that packets are transferred only from upstream node to downstream node in the Dedicated-VC. The packets from downstream node to upstream node are transferred through the Default-VC or through another Dedicated-VC.

3.2.2 VCID Notification Procedure

After the Dedicated-VC selection procedure, the upstream node transmits the PROPOSE message to the downstream node through the Dedicated-VC. The PROPOSE message contains a VCID for the Dedicated-VC and IP address (target IP address) of downstream node. When the downstream node accepts the PROPOSE message, it transmits the PROPOSE ACK message to the upstream node through the Default-VC. With this procedure, the upstream and the downstream nodes (both end-points of the Dedicated-VC) can share the same indicator "VCID" associated with the given Dedicated-VC. The upstream and downstream nodes establish the mapping between the VCID, that is common for both nodes, and the label (i.e., VPI/VCI value), that is only local significant for each (upstream / downstream) node interface. By the establishment of mapping between the VCID and local label (VPI/VCI) for each node interface, the egress label (VPI/VCI) value for upstream node and the ingress label (VPI/VCI) value for downstream node can be different values.

When the downstream node can not accept the proposal from the upstream node with some reason (e.g., policy), the downstream node sends an ERROR message to the upstream node through the Default-VC. The upstream node retransmits the PROPOSE message when it neither receive PROPOSE ACK message nor ERROR message. When the upstream node has received neither of the messages even with given number (five in this paper) of retransmissions of the PROPOSE message, the Dedicated-VC picked up through the Dedicated-VC selection procedure should be released.

The purpose of the VCID notification procedure is not only to share the VCID information regarding the Dedicated-VC, but also to confirm whether the Dedicated-VC is available and whether the neighbor node operates correctly. As discussed above, with VCID notification procedure, the label switching paradigm can be applied to the label swapped data-link (e.g., ATM links). In other words, with the VCID notification procedure, the ingress and the egress labels for a data-link channel used for label switching can be different values.

3.2.3 Flow-ID Notification Procedure

After the VCID notification procedure, the upstream node transmits an OFFER message to the downstream node through the Default-VC. The OFFER message contains the VCID of the Dedicated-VC, the flow-ID of the packet flow transferred through the Dedicated-VC and the refresh interval of a READY message.

When the downstream node receives the OFFER message from the upstream node, it transmits the READY message to the upstream node through the Default-VC in order to indicate that the OFFER message issued by the upstream node is accepted. By the reception of the READY message, the upstream node realizes that the downstream node can receive IP packets transferred through the Dedicated-VC.

The upstream node retransmits the OFFER message when it does not receive a READY message from the downstream node. When the upstream node has not receive a READY message even with given number (five in this paper) of retransmissions, the Dedicated-VC should be released.

When the downstream node accepts the OFFER message from the upstream node, it must send a READY message to the upstream node within the refresh interval offered by the upstream node. If it can not, the downstream node sends the ERROR message (this refresh interval is not supported) to the upstream node. The downstream node should accept the refresh interval larger than the predefined period. Therefore the downstream node shouldn't send the ERROR message (this refresh interval is not supported) when the refresh interval in the OFFER message is larger than the predefined period.

By the Flow-ID notification procedure, the node establishes the mapping between the VCID and the corresponding Flow-ID. By the VCID notification procedure (i.e., mapping between VPI/VCI and VCID) and the Flow-ID notification procedure (i.e., mapping between VCID and Flow-ID), the node can establishes the mapping between the packet flow (i.e., Flow-ID) and the corresponding data-link label (i.e., VPI/VCI value).

3.2.4 Flow ID Refresh Procedure

While the downstream node receives IP packets through the Dedicated-VC, it should periodically (with a refresh interval) send the READY message to the upstream node. When the downstream node does not receive any IP packet during the refresh interval, it does not send the READY message to the upstream node.

While the upstream node continues to receive READY messages, it realizes that it can transmit the IP packets through the Dedicated-VC. When it does not receive a READY message at all for a predetermined period (dead interval), it removes the mapping between the Flow IP and VCID. The dead interval is defined below.

When the upstream node falls into failure without the Flow ID removal procedure for a Dedicated-VC, its mapping must be removed by the downstream node. The downstream node removes the mapping between the Flow ID and VCID for the Dedicated-VC when it does not receive any IP packet for a "removal period" (=refresh interval times m). The refresh interval, the dead interval and the removal period should satisfy the following equation.

$$\begin{aligned} &refresh\_interval < dead\_interval \\ &\quad < removal\_period (= refresh\_interval \times m) \end{aligned}$$

3.2.5 Flow ID Removal Procedure

When the upstream node realizes that the Dedicated-VC is not used, it performs a Flow ID removal procedure. The Flow ID removal procedure differs between the case of PVC/VP configuration and the case of SVC configuration. With the PVC/VP configuration, the upstream node issues a REMOVE message to the downstream node, and the downstream node sends back a REMOVE ACK message to the upstream node. The upstream node retransmits REMOVE messages when it does not receive a REMOVE ACK message. The upstream node assumes that the downstream node is in failure state when it dose not receive any REMOVE ACK message from the downstream node even with five REMOVE message retransmissions.

With SVC configuration, two procedures are possible. One is that the mapping between the Flow ID and the VCID is removed without the release of the ATM connection, which is the same procedure as the PVC/VP configuration. The other procedure is that the mapping between the Flow ID and the VCID is removed by releasing the VC through ATM signaling. The former procedure can promptly create and delete the mapping between Flow ID and VCID, since the ATM signaling does not have to be performed each time. However, an un-used ATM connections have to be maintained by the node. Which procedure is applied to is a matter of each CSR's local decision, taking the VC resource cost and responsiveness into account.

The downstream node may want to remove the mapping between the Flow ID and the VCID. When the upstream node receives the REMOVE message, it sends a REMOVE ACK message to the downstream node.

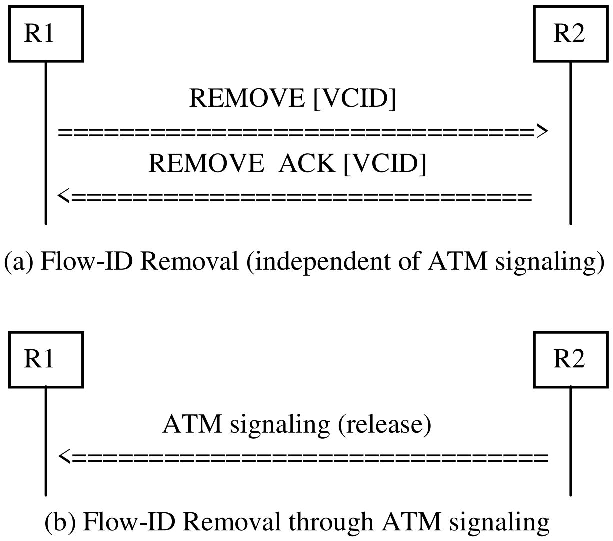

Figure 3. Flow ID Removal Procedure

4. Performance and Feasibility Evaluation

4.1 Interworking with ATM Interface

With the VCID notification and Flow-ID notification procedures described in section 3 and in [6], the label switching router applying FANP can work over the standard ATM link. However, the routers applying the other label management protocol, e.g., TDP, IFMP, can not work over the standard ATM link. This means that the router applying non-FANP can not use the standard ATM links to interconnect them and can not interconnect with the standard ATM switch. On the contrary, the label switch routers applying FANP can use the standard ATM links to interconnect them, and can use standard ATM switches as the data-link switching hub.

4.2 Cut-through Path Establishment Delay

4.2.1 Evaluation Model

The following is the configuration of CSR prototype system with FANP supporting SVC.

- CSR IP controller module;
  - CPU; Pentium 166MHz
  - OS; BSD/OS v2.1
  - NIC card; Efficient Network 155 PCI NIC with 2MB memory
- CSR cell-switch module; FORE ASX-200WG
- CSR cell-switch module control; SNMP through Ethernet interface
- Cut-through Parameters for Flow Driven mode(configurable)
  - Applications to be cut-through: web, ftp, telnet, nntp
  - Cut-through time out : two minutes

Figure 4 shows the overview of the evaluation system using SVC and PVC connections to transfer the IP packets. Two ATM switches are interconnected through a CSR. Each ATM switch has it's own IP subnet. Each ATM switch (AX-1500) accommodate the host, that applies FANP.

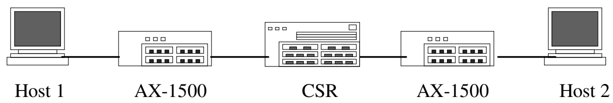

Figure 4. Configuration of performance evaluation

4.2.2 Evaluation Result

Table 1 gives the average cut-through path establishment delay with FANP, when a new cut-through path is established. In the evaluation, there is no IP packet flow before the cut-through path is established. The most contribution regarding the cut-through path establishment using SVC is signaling delay between CSRs and cell switch module configuration delay in the CSR. 88 msec in 95 msec is required by the SVC operation, i.e., about ten cut-through paths establishment per second. Regarding PVC operation, the delay for cut-through path establishment is 7 msec, i.e., more than 100 cut-through paths establishment per second. In the actual system, the cut-through path establishment delay will be larger than this evaluation. This is because there is some tasks of handling previously existing cut-through path and hop-by-hop packet flows. The further detailed evaluation should be for further study item. The major factor to increase the cut-through establishment delay would be the amount of hop-by-hop packet flows.

4.2.3 Cut-through Path Establishment Delay Improvement

As evaluated in the previous section, the delay to establish the cut-through path will be sometimes more than

Table 1. Cut-through path establishment delay with FANP

	Delay
Signaling Between CSRs	58 msec.
Cell Switch Module Configuration	30 msec.
VCID Notification	3 msec.
Flow Notification	4 msec.

100 msec. In order to improve the delay performance to establish cut-through path, the following mechanism will be introduced. Some unused VCs (called as stand-by Dedicated VC) for the future use as the Dedicated VC are established, in advance. When the trigger packet to establish the cut-through path appears at the CSR, one of stand-by Dedicated VCs is picked up as the Dedicated VC to be used for the appeared packet flow. After one stand-by Dedicated VC is picked up for the actual cut-through packet forwarding, a new stand-by Dedicated VC will be established. Through this operation, the SVC signaling delay across the ATM switch(es), that should be major delay factor for cut-through path establishment, can be dramatically reduced.

The other method to reduce the cut-through path establishment delay performance would be the reduce of Dedicated VC configuration in the cell-switch module in CSR. Special command or instruction could be defined to configure the Dedicated VC in the cell-switch module.

Those method should be evaluated as a for further study items with a future implementation.

4.3 Effectiveness of Cut-through Packet Forwarding

Figures 5-7 give the effectiveness of cut-through packet forwarding, e.g., percentage of packets forwarded with cut-through path. Figures are regarding with the actual packet flow data of the gateway router from the Digital Equipment Corporation network toward the Internet [11] in March of 1995. The data is for actual user traffic.

The percentage of packets for major applications are as followed;

1. web ; 13.41 %
2. telnet ; 0.43 %
3. ftp (file transfer protocol) ; 32.91 %
4. nntp (network news transfer protocol) ; 39.94 %
5. other applications ; 13.31 %

With larger refresh timer, the percentage of packets forwarded through cut-through paths will be larger,

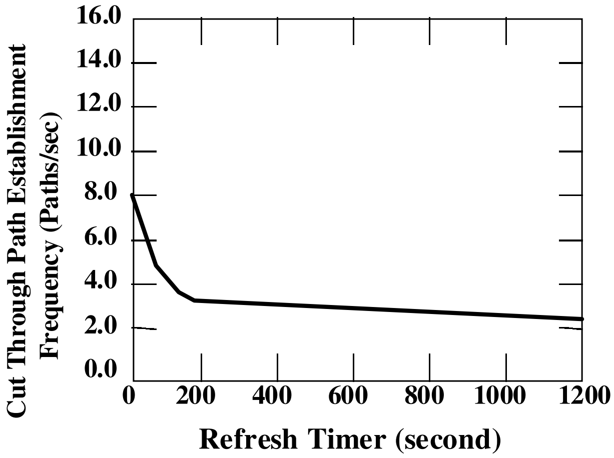

Figure 5. Cut-through Path Establishment Frequency Evaluation

but the required number of cut-through paths to maintain by the CSR becomes larger. With smaller refresh timer, the percentage of packets forwarded through cut-through paths may reduce and the required cut-through path establishment frequency will become larger, however the required number of cut-through paths to maintain by CSR becomes smaller.

As shown in figure 5, the required cut-through path establishment performance will be 10 per second, even with 10 seconds refresh interval to manage the time out to release the cut-through path. Figure 6 shows that the number of cut-through paths established in the CSR will be 1,200 paths and 1,000 paths in average, with 200 second of refresh interval. Finally, as shown in figure 7, the percentage of packets and data to be forwarded through the cut-through path will be more than 95%, regardless of the refresh interval.

These result shows that the refresh interval could be less than 100 seconds. This is by the following two reasons. (1) Even for small refresh interval, the required cut-through path establishment frequency will not significantly increase. (2) The percentage of packets to be forwarded through cut-through paths does not significantly decrease, even for small refresh interval. Therefore, in this example, the refresh timer could be less than 100 seconds. With this refresh interval, the required number of cut-through paths for this evaluation model will be less than 500.

5. Conclusion

This paper describes and performs the primary feasibility evaluation on the FANP (Flow Attribute Notification Protocol) to apply the label switching paradigm in CSR (Cell Switch Router) system.

FANP defines two key procedures, i.e., one is the

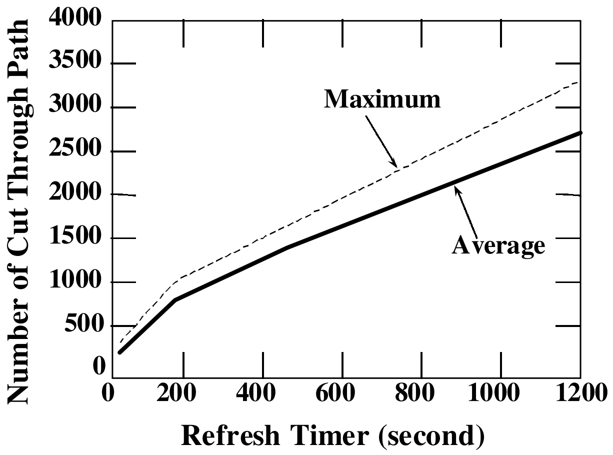

Figure 6. Number of Established Cut-through Paths

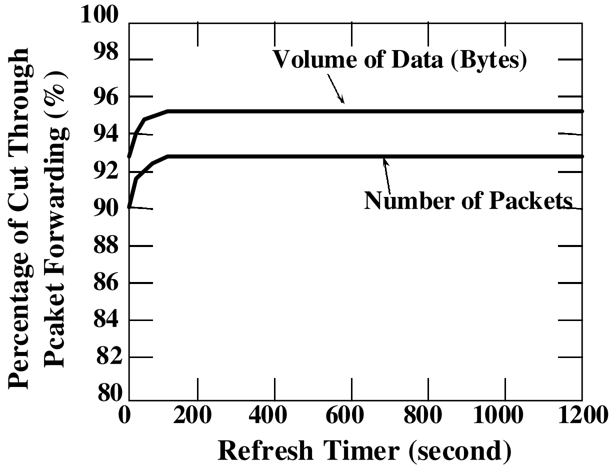

Figure 7. Percentage of Cut-through Packet Forwarding

VCID Notification Procedure, and the other is the Flow-ID Notification Procedure. The VCID Notification Procedure lets the label switching paradigm over the label swapped data-link, such as ATM links, though the other label switch architectures can not work over the label swapped data-link.

The primary evaluation results, using the prototype implementation and using the actual packet transmission date, show that the label switch router system with FANP would work well for corporate backbone level. In the Internet backbone level, it is expected that the required number of cut-through paths will significantly increase, and that we should introduce the flow aggregation based on network topology and network addresses. The introduction of the flow aggregation policy into FANP and its evaluation should be for further study item.

References

[1] R.Woundy, A.Viswanathan, N.Feldman, R.Boivie : "ARIS: Aggregate Route- Based IP Switching", IETF Internet-Draft, draft-woundy-aris-ipswitching-00.txt, November, 1996.

[2] P.Newman, T.Lyan, G.Minshall : "Flow labelled: Connectionless ATM Under IP", Engineer Conference, Networld+Interop'96 Las Vegas, April, 1996.

[3] Y.Katusbe, K.Nagami, H.Esaki : "Toshiba's Router Architecture Extensions for ATM : Overview", IETF RFC2098, Feb., 1997.

[4] Y.Rekhter,B.Davie,D.Katz,E.Rosen,G.Swallow : "Cisco Systems's Tag Switching Architecture Overview", IETF RFC2105, Feb., 1997.

[5] P.Newman, et.al, : "Ipsilon Flow Management Protocol Specification for IPv4 - version 1.0 -", IETF RFC1953, May, 1996.

[6] K.Nagami, et.al, "Toshiba's Flow Attribute Notification Protocol (FANP) Specification", IETF RFC2129, April, 1997.

[7] P.Doolan, B.Davie, D.Katz, Y.Rekhter, E.Rosen : "Tag Distribution Protocol", IETF Internet-Draft, draft-doolan-tdp-spec-00.txt, September, 1996.

[8] H.Esaki, M.Ohta, K.Nagami : "High Speed Datagram Delivery over Internet using ATM Technology", IEICE Transactions on Communications, Vol.E78-B No.8, August 1995

[9] H.Esaki, M.Ohta, K.Nagami : "High Speed Packet Delivery over Internet using ATM Technology", Engineer Conference, Networld+Interop95 Las Vegas, March, 1995.

[10] Y.Katsube, K.Nagami, H.Esaki : "Cell Switch Router - Basic Concept and Migration Scenario", Engineer Conference, Networld+Interop96 Las Vegas, April, 1996.

[11] http://ita.ee.lbl.gov/html/contrib/DEC-PKT.html, March 1995.

[12] Juha Heinanen, "Multiprotocol Encapsulation over ATM Adaptation Layer 5", IETF RFC1483,July.,1993.

[13] M.Laubach : "Classical IP and ARP over ATM", IETF RFC1577, October 1993.

Appendix A: Message Formats

FANP control procedure includes seven messages described in section 3. Among them, a PROPOSE message used for VCID notification procedure uses an extended ATM ARP message format defined in RFC1577 [13]. The other messages are encapsulated into IP packets. The destination IP address in the IP packet header signifies the neighbor node's IP address and the source IP address signifies sender's IP address. Currently, the protocol ID for these messages is 110(decimal). This protocol ID must be registered by IANA. The reserved field in the following packet format must be zero.

Appendix A:1 Field Format

Appendix A:1.1 VCID field

VCID type value decides VCID field format. Currently, only type "1" is defined. The VCID field format of VCID type 1 is shown below.

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                     ESI of upstream node                      |
+                               +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                               |                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+                               +
|                               ID                              |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

ESI field: ESI of upstream node

ID : upstream node decides unique identifier.

Appendix A:1.2 Flow ID field

Flow ID type value decides flow-ID field format. Currently, flow-ID type "0" and "1" are defined. The flow ID type value "0" signifies that the flow ID field is null. When flow ID type value is "1", the format shown below is used.

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                       Source IP address                       |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                     Destination IP address                    |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Source IP address : source IP address of flow

Destination IP address : destination IP address of flow

Appendix A:2 PROPOSE message

PROPOSE message uses the extended ATM-ARP message format [12] to which the VCID type and the VCID field are added. Type & Length fields are set to zero, because the messages don't need sender/target ATM address. This message is transferred from the upstream node to the downstream node through the Dedicated-VC. PROPOSE message is transferred from the upstream node to the downstream node through the Dedicated-VC.

Appendix A:3 PROPOSE ACK

PROPOSE ACK messages is transferred through the Default-VC.

Appendix A:4 OFFER message

OFFER message is transferred from an upstream node to a downstream node. The following is the message format.

Appendix A:5 READY message

READY message is transfered from a downstream node to an upstream node. This message is transferred when the downstream node receives OFFER message. And

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| Hardware Type = 0x13          |  Protocol Type = 0x0800       |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| Type&Length 1 | Type&Length 2 |      Operation Code           |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|    Length 1   | Type&Length 3 | Type&Length 4 |   Length 2    |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                       Sender IP Address                       |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                       Target IP Address                       |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|   VCID Type   |VCID Length    |          Reserved             |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                         VCID                                  |
/                                                               /
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Type&Length 1 ; Type & Length of sender ATM number = 0

Type&Length 2 ; Type & Length of sender ATM subnumber = 0

Type&Length 3 ; Type & Length of sender ATM number = 0

Type&Length 4 ; Type & Length of sender ATM subnumber =0

Length 1 ; Source IP address length

Length 2 ; Target IP address length

Operation code

0x10 = PROPOSE

VCID Type: Currently , VCID Type = 1 is defined.

VCID Length: Length of VCID field

VCID : VCID described previous

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| Version       |Op code = 1    |          Checksum             |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| VCID type     |Flow-ID type=0 |          Reserved             |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                             VCID                              |
/                                                               /
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Version

This field indicates the version number of FANP. Currently, Version = 1

Operation Code

This field indicates the operation code of the message. There are five operation codes, below.

operation code = 1 : PROPOSE ACK message

Checksum

This filed is the 16 bits checksum for whole body of FANP message. The checksum algorithm is same as the IP header.

VCID Type

This field indicates the VCID type. Currently, only "1" is defined.

this message is transferred periodically in each refresh interval. The following is the message format.

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| Version = 1   | Op Code = 2   |           Checksum            |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| VCID type     |Flow-ID type   |       Refresh Interval        |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                              VCID                             |
/                                                               /
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                            Flow-ID                            |
/                                                               /
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Refresh Interval

This field indicates the interval of refresh timer. The refresh interval is represented by second in integer. This field is used only in OFFER message.

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| Version = 1   | Op Code = 3   |           Checksum            |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| VCID type     |Flow-ID type   |       Reserved                |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                              VCID                             |
/                                                               /
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                            Flow-ID                            |
/                                                               /
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Appendix A:6 ERROR message

ERROR message is transferred from a downstream node to an upstream node or from an upstream node to a downstream node. This message is transferred when some of the fields in the receive message is unknown or refused. When the receive message is the ERROR message, ERROR message isn't sent. VCID type ,VCID, Flow ID Type and Flow ID field in the ERROR message are filled with the same field in the receive message. The following is the message format.

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| Version = 1   | Op Code = 4   |           Checksum            |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| VCID type     |Flow-ID type   |       Error code              |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                              VCID                             |
/                                                               /
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                            Flow-ID                            |
/                                                               /
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Error Code = 1 : unknown VCID type
= 2 : unknown Flow-ID type
= 3 : unknown VCID
= 4 : resource is unavailable
= 5 : unavailable refresh interval is offered
= 6 : refuse by policy

Appendix A:7 REMOVE message

REMOVE message is transferred from a downstream node to an upstream node or vice versa. This message is transferred to remove the mapping relationship between the flow ID and the VCID. The node which receives REMOVE message must send REMOVE ACK message, even when VCID in the receive message isn't known.

The following is the message format.

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| Version = 1   | Op Code = 5   |           Checksum            |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| VCID type     |Flow-ID type   |       Reserved                |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                              VCID                             |
/                                                               /
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Appendix A:8 REMOVE ACK message

REMOVE ACK message is transferred from a downstream node to an upstream node or from an upstream node to a downstream node. The following is the message format.

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| Version = 1   | Op Code = 6   |           Checksum            |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| VCID type     |Flow-ID type   |       Reserved                |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                              VCID                             |
/                                                               /
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Appendix B: Terminology and Definition

1. VCID (Virtual Connection IDentifier)
   Since VPI/VCI values at the origination and the termination points of a VC (and VP) may not be the same, we need an identifier to uniquely identify the data-link connection between neighbor nodes. We define this identifier as a VCID. Currently, only one type of VCID is defined in FANP. This VCID contains the ESI (End System Identifier) of a source node and the unique identifier within a source node.

2. Flow ID (Flow IDentifier)
   IP level packet flow is identified by some parameters in a packet. Currently, only one type of flow ID is defined in FANP. This flow ID contains a source IP address and a destination IP address. Note that flow ID used in this paper is not the same as the flow-id specified in IPv6.

3. Cut-through packet forwarding
   Packets are forwarded without any IP processing at the router using the data-link level information (e.g., VPI/VCI). Internetworking level information (e.g., destination IP address) is mapped to the corresponding data-link-level identifier by using the FANP.

4. Hop-by-Hop packet forwarding
   Packets are forwarded using IP level information like conventional routers. In ATM, cells are reassembled into packets at the router to analyze the IP header.

5. Default-VC
   Default-VC is used for hop-by-hop packet forwarding. Cells received from the Default-VC are reassembled into IP packets. Conventional IP processing is performed for these packets. The encapsulation over the Default-VC is LLC for routed non-ISO defined by RFC1483 [12], in FANP.

6. Dedicated-VC
   Dedicated-VC is used for the specific IP packet flow identified by the flow-ID. When the flow-ID for an incoming VC and an outgoing VC are the same at a CSR, it can forward the packets belonging to the flow through the cut-through packet forwarding. The encapsulation over the Dedicated-VC is LLC for routed non-ISO protocols defined by RFC1483 [12], in FANP.

7. Cut-through trigger
   When a FANP capable node receives a trigger packet, it tries to establish Dedicated-VC and to notify the mapping information between the Dedicated-VC and the IP packet flow which the received trigger packet belongs to. Trigger packets are defined by the port-ID of TCP/UDP with the local policy of each FANP capable node. In general, they would be the port-ID's of sessions with a long life-time and/or with large amount of packets; e.g., http, ftp and nntp. Future implementation will include other triggers such as an arrival of resource reservation request.

8. Label swapping data-link
   In the label swapping data-link, such as the ATM link, the data-link label (e.g., VPI/VCI value in the ATM link) is swapped at every data-link node. Therefore, the ingress and egress data-link labels are generally the different value for given data-link virtual connection (e.g., ATM-VC/VP).

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

Yasuhiro Katsube received the B.E. and M.E degrees in electrical engineering from Keio University, Yokohama, Japan, in 1984 and 1986, respectively. In 1986, he joined Toshiba Research and Development Center, Kawasaki, Japan, where he has been engaged in the research and development of ATM network systems. From 1992 to 1994, he was a visiting researcher at Bell Communications Research. He is currently at the Communication and Information Systems Research Labs, Toshiba Research and Development Center.

Yasuro Shobatake received B.E., M.E. and Ph.D from Keio University in 1985, 1987, 1995, respectively. In 1987, he joined Research and Development Center, Toshiba Corp., where he has been engaged in the research and development of switching technologies, mainly ATM. From 1996 to 1998 he was a visiting researcher at Center for Telecommunications Research, Columbia University. He has received 1993 Young Engineer Award from IEICE. He is a member of Information Processing Society of Japan and ACM.

Akiyoshi Mogi received the B.E. degree from Waseda University, Japan, in 1989 and M.E.degree from Waseda University, Japan, in 1991. He joined Research and Development Center, Toshiba Corporation in 1991. He has been working on computer system software such as compiler and operating system. He has been working with Cell Switch Router project from 1995.

Shigeo Matsuzawa received the B.E. degree from Waseda University, Japan, in 1990. 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.

Tatsuya Jinmei received the B.S. degree from Kyoto University, Japan, in 1993 and M.E. degree from Kyoto University, Japan, in 1995. He joined Research and Development Center, Toshiba Corporation where he has been working on communication system. He is currently interested in internet technology for high speed networks.

Hiroshi Esaki received the B.E. and M.E. degrees from Kyushu University, Fukuoka, Japan, in 1985 and 1987, respectively. And, he received Ph.D from University of Tokyo, Japan, in 1998. 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 a high speed internet architeture and a mobile computing.