Network Working Group
Request for Comments: 2129
Category: Informational

K. Nagami
Y. Katsube
Y. Shobatake
A. Mogi
S. Matsuzawa
T. Jinmei
H. Esaki
Toshiba R&D Center

April 1997

Toshiba's Flow Attribute Notification Protocol (FANP) Specification

Status of this Memo

Abstract

This memo discusses Flow Attribute Notification Protocol (FANP), which is a protocol between neighbor nodes for the management of cut-through packet forwarding functionalities. In cut-through packet forwarding, a router doesn't have to perform conventional IP packet processing for received packets. FANP indicates mapping information between a datalink connection and a packet flow to the neighbor node and helps a pair of nodes manage the mapping information. By using FANP, routers (e.g., CSR; Cell Switch Router) can forward incoming packets based on their datalink-level connection identifiers, bypassing usual IP packet processing. The design policy of the FANP is;

(1) soft-state cut-through path (Dedicated-VC) management
(2) protocol between neighbor nodes instead of end-to-end
(3) applicable to any connection oriented datalink platform

1. Background

Due to the scalability requirement, connection oriented (CO) datalink platforms, e.g., ATM and Frame Relay, are going to be used as well as connection less (CL) datalink platforms, e.g., Ethernet and FDDI. One of the important features of the CO datalink is the presence of a datalink-level connection identifier. In the CO datalink, 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. CSRs [1] 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, CSR determines the egress interface and egress VPI/VCI value.

In order to configure the cut-through packet forwarding state, a pair of neighbor nodes have to share the mapping information between the packet flow and the datalink VC. FANP (Flow Attribute Notification Protocol) described in this memo is the protocol to configure and manage the cut-through packet forwarding state.

2. Protocol Requirements and Future Enhancement

2.1 Protocol Requirements

The followings are the protocol requirements for FANP.

(1) Applicable to various types of CO datalink platforms

(2) Available with various connection types (i.e., SVC, PVC, VP)

(3) Robust operation
The system should operate correctly even under the following conditions.

(a) VC failure
Some systems can detect VC failure as the function of datalink (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.

(b) Message loss
Control messages in the FANP may be lost. The system has to operate correctly, even when some control messages are lost.

(c 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, the following discussion assumes that the datalink is an ATM network.

2.2 Future Enhancement

The followings are the future enhancements to be done.

(1) Aggregated flow

In this memo, we define the flow which contain source and destination IP address. As this may require many VC resources, we also need a new definition of aggregated flow which includes several end-to-end flows. The concrete definition of the aggregated flow is for future study.

(2) Providing multicast service
(3) Supporting IP level QOS signaling like RSVP
(4) Supporting IPv6

3. Terminology and Definition

- 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 datalink connection between neighbor nodes. We define this identifier as a VCID. Currently, only one type of VCID is defined. This VCID contains the ESI (End System Identifier) of a source node and the unique identifier within a source node.

- 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. This flow ID contains a source IP address and a destination IP address. Note that flow ID used in this specification is not the same as the flow-id specified in IPv6.

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

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

- 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 protocols defined by RFC1483 [3].

- 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 [3].

- 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.

4. 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 between the neighboring nodes only. 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.

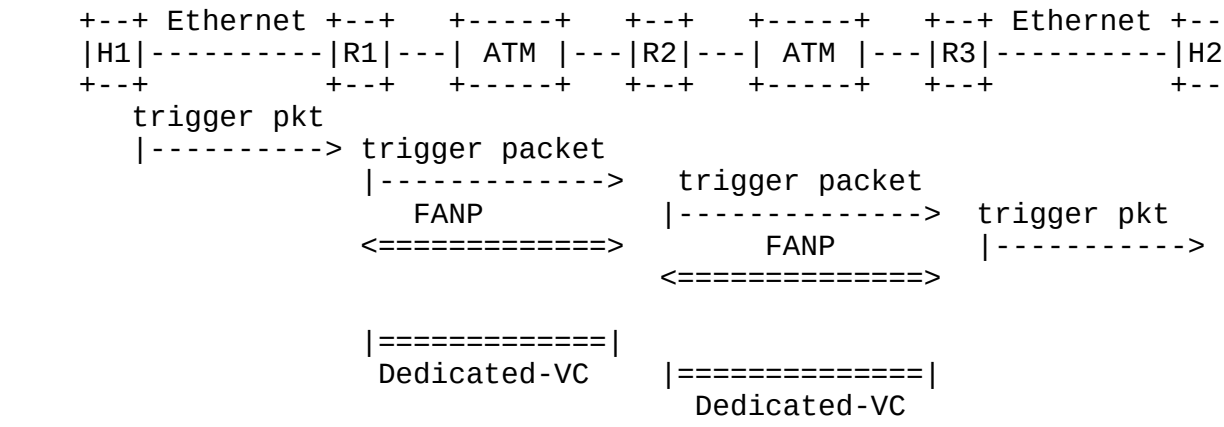


Figure 1. Trigger packet and FANP initiation

5. Protocol Sequence

FANP has the following five procedures, that are (1) Dedicated-VC selection, (2) VCID negotiation, (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.

5.1 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 un-used 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 used, a Dedicated-VC is established through ATM signaling each time it is required.

The procedure (1) can decrease a time to activate a Dedicated-VC. But the necessary VC resource will increase as it need to prepare additional VCs. Which procedure should be applied to is a matter of local decision in each node, taking the economical requirement and the system responsiveness into account.

A Dedicated-VC is used as a uni-directional VC, although it is generally bi-directional. 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.

5.2 VCID Negotiation 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" for the Dedicated-VC. 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 procedure at the downstream node which has received PROPOSE message is;

1. if(Target IP address of the PROPOSE message isn’t equal to my IP address)
   then Goto end.

2. if(The PROPOSE message should be refused)
   then Send an ERROR(refuse by policy) message. Go to end.

3. if(VCID Type in the PROPOSE message isn’t known)
   then Send an ERROR(unknown VCID Type) message. Go to end.

4. if(The VCID in the PROPOSE message is the same as the VCID which has already been registered for another Dedicated-VC in the node)
   then Delete the registered VCID.
   Release the old Dedicated-VC.

5. if(A VCID is registered for the Dedicated-VC which has received the PROPOSE message)
   then Delete the registered VCID.

6. Register the mapping between VCID and I/F, VPI, VCI for the Dedicated-VC.

7. if(The mapping is successful)
   then Send a PROPOSE ACK.
   else Send an ERROR(resource unavailable).

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 five retransmissions of the PROPOSE message, the Dedicated-VC picked up through the Dedicated-VC selection procedure should be released. Here, the number of retransmissions (five in this specification)is recommended value and can be modified in the future.

The purpose of the VCID negotiation 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.

If the VCID negotiation procedure with a neighbor node always fails, it is considered that the node may not be FANP-capable node. Therefore the upstream node should not try the VCID negotiation procedure to that node for a certain time period.

5.3 Flow-ID Notification Procedure

After the VCID negotiation 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 five retransmissions, the Dedicated-VC should be released. Here, the number of retransmissions (i.e., five in this specification) is a recommended value and may be modified in the future.

The node transmits an ERROR message to its neighbor in the following cases. When the node receives the ERROR message, the Dedicated-VC should be released.

(a) unknown VCID: The VCID in the message is unknown.
(b) unknown VCID Type: The VCID Type is unknown.
(c) unknown flow-ID Type: the flow-ID Type is unknown.

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 120 seconds. 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 120 seconds.

The following describes the procedure of the node which has received an OFFER message.

1. if(unknown version in the OFFER message)
   then Discard the message. Goto end.

2. if(unknown VCID Type in the OFFER message)
   then Send an ERROR (unknown VCID Type) message. Goto end.

3. if(VCID in the OFFER message has not been registered)
   then Send an ERROR (unknown VCID) message. Goto end.

4. if(unknown Flow ID Type in the OFFER message)
   then Send an ERROR (unknown Flow ID Type) message. Goto end.

5. if(refuse Flow ID in the OFFER message)
   then Send an ERROR (refused by policy) message. Goto end.

6. if(refuse refresh interval in the OFFER message)
   then Send an ERROR(This refresh interval is not supported)
   message. Goto end.

7. if(the mapping between Flow ID and VCID already exists and
   Flow ID in the OFFER message is different from the registered
   Flow ID for the corresponding VCID)
   then Do Flow-ID removal procedure. Goto end.

8. Do the procedure of receiving the OFFER message.

7. if(successful)
   then Send a READY message.
   else Send an ERROR (resource unavailable) message.

8. end.

The procedure of the node which has received a READY message is described.

1. if(unknown version in the READY message)
   then Discard the message. Goto end.

2. if(unknown VCID Type in the READY message)
   then Send an ERROR (unknown VCID Type) message. Goto end.

3. if(VCID in the READY message has not been registered)
   then Send an ERROR (unknown VCID) message. Goto end.

4. if(unknown Flow ID Type in the READY message)
   then Send an ERROR (unknown Flow ID Type) message. Goto end.

5. if((the mapping between Flow ID and VCID doesn't exist)||
   (the mapping between Flow ID and VCID already exists and
   Flow ID in the READY message is different from registered Flow
   ID for the corresponding VCID))
   then Send an ERROR (unknown VCID) message. Goto end.

6. Do the procedure of receiving the READY message.

7. end.

5.4 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.

refresh interval < dead interval < removal period (=refresh interval times m)

The recommended values are:

```
refresh interval =  2 minutes
dead interval    =  6 minutes (=refresh interval x 3)
removal period   = 20 minutes (=refresh interval x 10)
```

5.5 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.

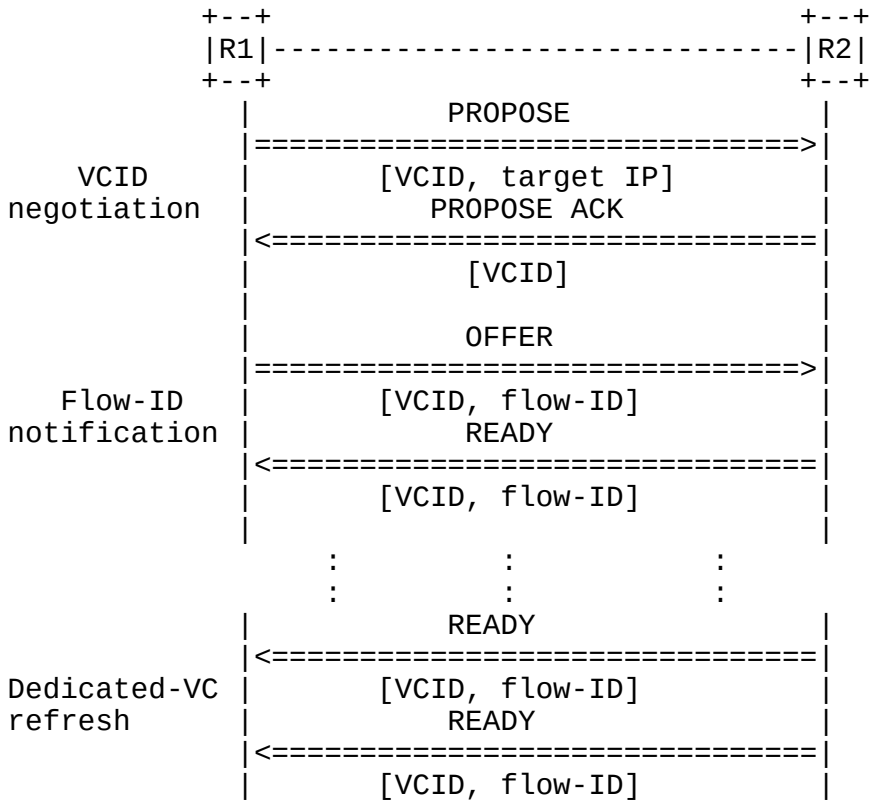
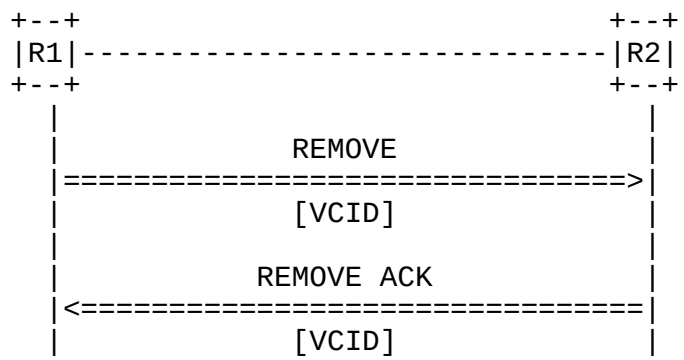


Figure 2. Flow ID notification and refresh procedure



(a) Flow ID removal (independent of ATM signaling)

(b) Flow ID removal through ATM signaling

Figure 3. Flow ID removal procedure

6. Message Format

FANP control procedure includes seven messages described from 6.2 to 6.8. Among them, a PROPOSE message used for VCID negotiation procedure uses an extended ATM ARP message format defined in RFC1577 [2]. The other messages are encapsulated into IP packets.

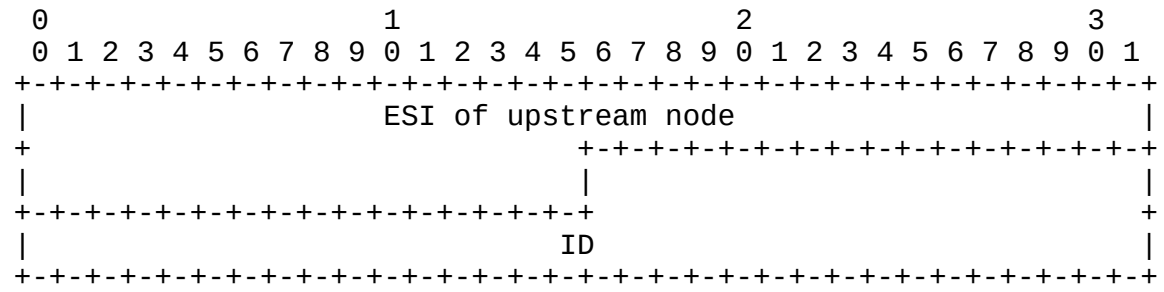
The destination IP address in the IP packet header signifies the neighbor node's IP address and the source IP address signifies sender's IP address. Currently, the protocol ID for these messages is 110(decimal). This protocol ID must be registered by IANA.

The reserved field in the following packet format must be zero.

6.1 Field Format

6.1.1 VCID field

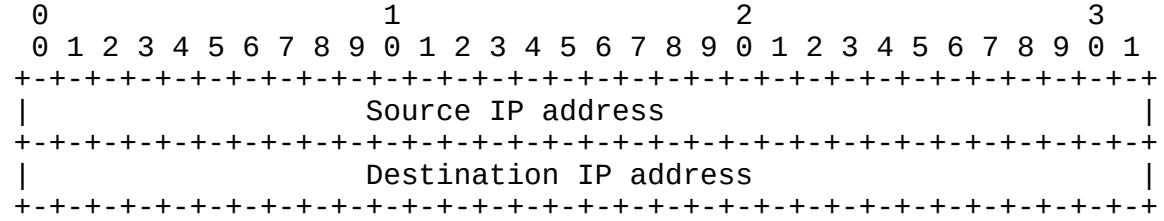
VCID type value decides VCID field format. Currently, only type "1" is defined. The VCID field format of VCID type 1 is shown below.



ESI field: ESI of upstream node
ID : upstream node decides unique identifier.

6.1.2 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.

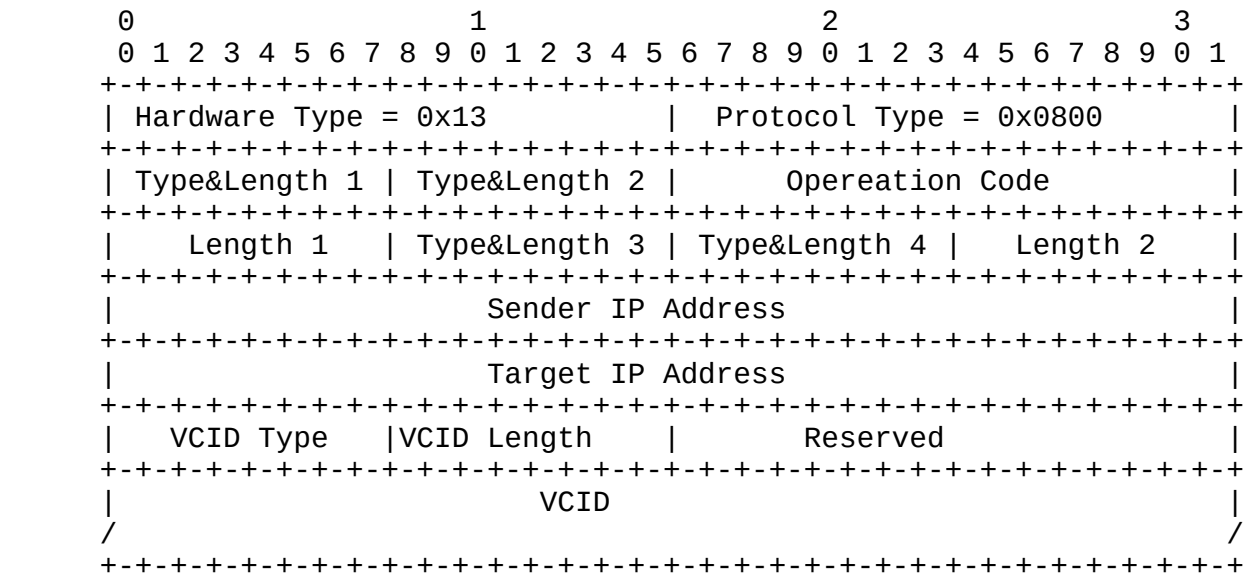


Source IP address : source IP address of flow
Destination IP address : destination IP address of flow

6.2 PROPOSE message

PROPOSE message uses the extended ATM-ARP message format [2] 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.



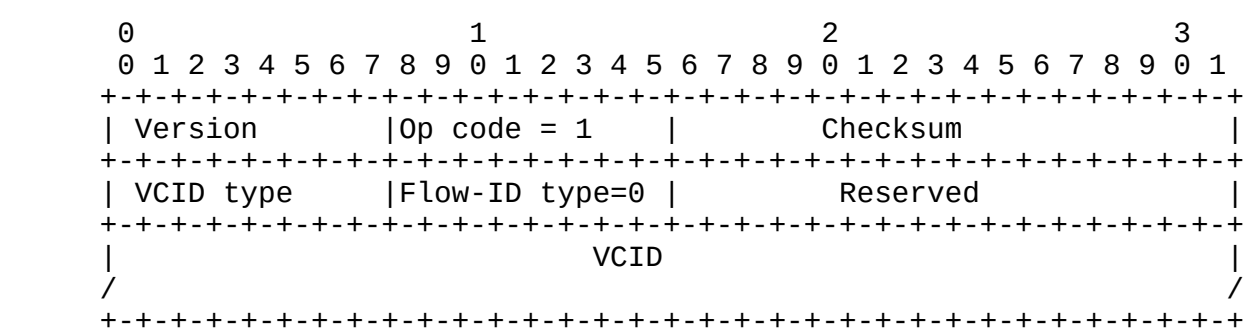
Type&Length 1 ; Type & Length of sender ATM number = 0
Type&Length 2 ; Type & Length of sender ATM subnumber = 0
Type&Length 3 ; Type & Length of sender ATM number = 0
Type&Length 4 ; Type & Length of sender ATM subnumber =0
Length 1 ; Source IP address length
Length 2 ; Target IP address length

Operation code
0x10 = PROPOSE

VCID Type: Currently , VCID Type = 1 is defined.
VCID Length: Length of VCID field
VCID : VCID described previous

6.3 PROPOSE ACK

PROPOSE ACK messages is transferred through the Default-VC.



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.

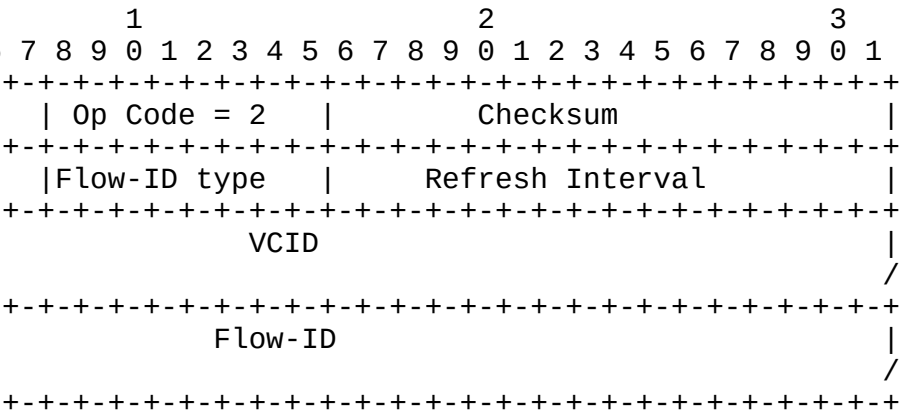
operation code = 1 : PROPOSE ACK message

Checksum
This field 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.

6.4 OFFER message

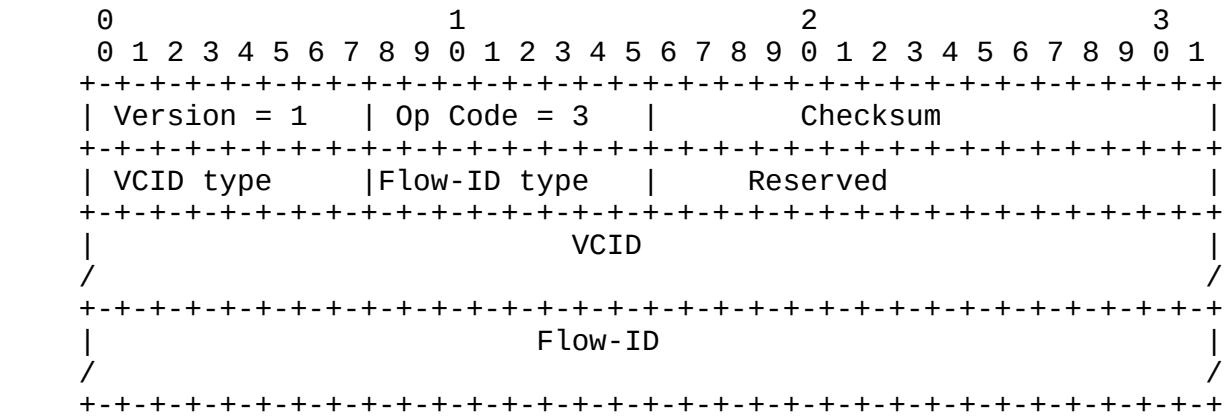
OFFER message is transferred from an upstream node to a downstream node. The following is the message format.



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. Recommended value is 120 (second).

6.5 READY message

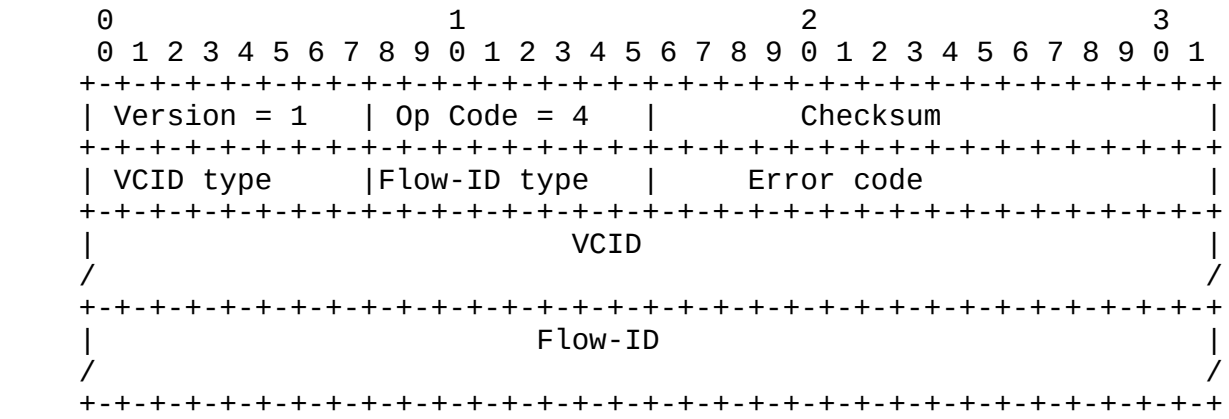
READY message is transfered from a downstream node to an upstream node. This message is transferred when the downstream node receives OFFER message. And this message is transferred periodically in each refresh interval. The following is the message format.



6.6 ERROR message

ERROR message is transfered 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.

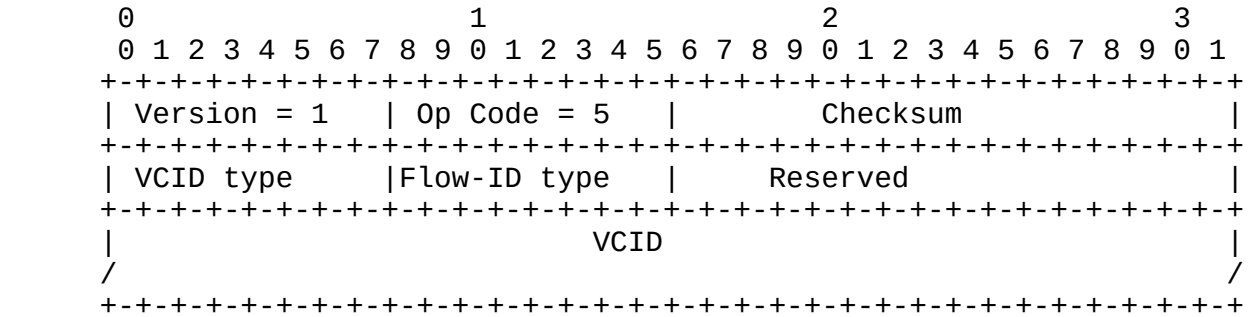


```
Error Code = 1 : unknown VCID type
           = 2 : unknown Flow-ID type
           = 3 : unknown VCID
           = 4 : resource is unavailable
           = 5 : unavailable refresh interval is offered
           = 6 : refuse by policy
```

6.7 REMOVE message

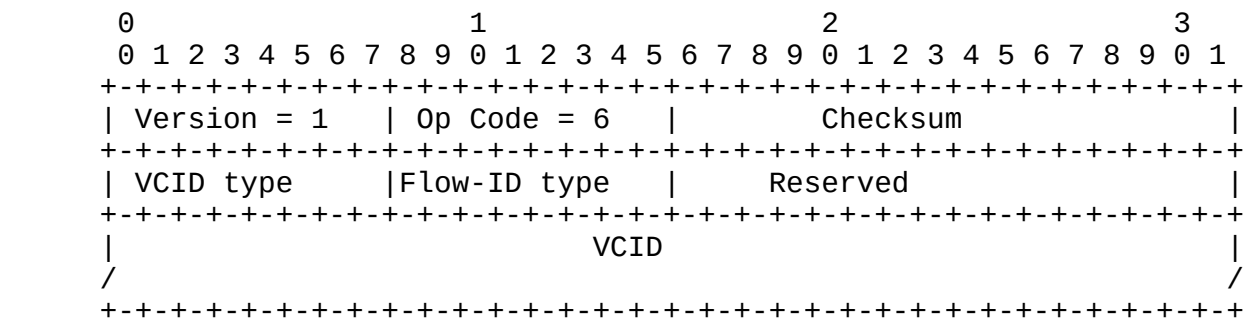
REMOVE message is transfered from a downstream node to an upstream node or vice versa. This message is transferred to remove the mapping relationship between the flow ID and and the VCID. The node which receives REMOVE message must send REMOVE ACK message, even when VCID in the receive message isn't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

The following is the message format.



6.8 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.



7. Security Considerations

Security issues are not discussed in this memo.

8. References

[1] Katsube, Y., Nagami, K., and H. Esaki, "Router Architecture Extensions for ATM; overview", Work in Progress.

[2] Laubach, M., "Classical IP and ARP over ATM", RFC 1577, October 1993.

[3] Heinanen, J., "Multiprotocol Encapsulation over ATM Adaptation Layer 5", RFC 1483, July 1993.

9. Authors' Addresses

Ken-ichi Nagami
R&D Center, Toshiba
1 Komukai Toshiba-cho, Saiwai-ku, Kawasaki 210 Japan
Phone : +81-44-549-2238
EMail : nagami@isl.rdc.toshiba.co.jp

Yasuhiro Katsube
R&D Center, Toshiba
1 Komukai Toshiba-cho, Saiwai-ku, Kawasaki 210 Japan
Phone : +81-44-549-2238
EMail : katsube@isl.rdc.toshiba.co.jp

Yasuro Shobatake
R&D Center, Toshiba
1 Komukai Toshiba-cho, Saiwai-ku, Kawasaki 210 Japan
Phone : +81-44-549-2238
Email : masahata@csl.rdc.toshiba.co.jp

Akiyoshi Mogi
R&D Center, Toshiba
1 Komukai Toshiba-cho, Saiwai-ku, Kawasaki 210 Japan
Phone : +81-44-549-2238
EMail : mogi@isl.rdc.toshiba.co.jp

Shigeo Matsuzawa
R&D Center, Toshiba
1 Komukai Toshiba-cho, Saiwai-ku, Kawasaki 210 Japan
Phone : +81-44-549-2238
EMail : shigeom@isl.rdc.toshiba.co.jp

Tatsuya Jinmei
R&D Center, Toshiba
1 Komukai Toshiba-cho, Saiwai-ku, Kawasaki 210 Japan
Phone : +81-44-549-2238
EMail : jinmei@isl.rdc.toshiba.co.jp

Hiroshi Esaki
R&D Center, Toshiba
1 Komukai Toshiba-cho, Saiwai-ku, Kawasaki 210 Japan
Phone : +81-44-549-2238
EMail : hiroshi@isl.rdc.toshiba.co.jp