IEEE 802.3 Link Aggregation

Flush Requirements
and
Operation

IBM
Jeff Lynch, Loren Larsen, Arush Kumar,
Mike Siegel
jjlynch@us.ibm.com

Austin, TX Meeting
9/1/98

Flush

Anytime an active flow(s) is moved from one physical link to another, a flush operation (or some other mechanism such as a time-out) is required to ensure that frames are not delivered out of order.

- Using an explicit Flush command rather than a time-out speeds things up

Potential Flush Scenarios

- **When a link is added to a LAG**

- **When a link is removed from a LAG**

- **When dynamically re-balancing active flows over a LAG**

- **. . .**

Flush Operation

- **A special control frame that is transmitted to and "echoed" back by the receiving station**

- **All traffic for <u>flows being affected</u> is held after a flush is sent until response is received**
  - **Ensures that the link is purged of relevant data**
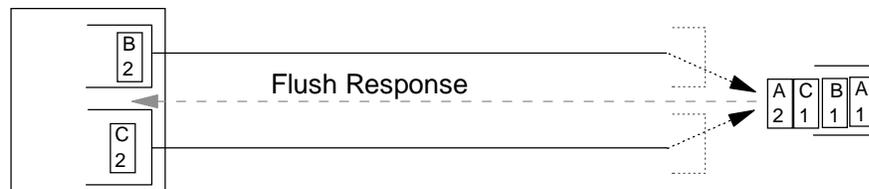 - **Other flows can continue to be sent**

Flush command

Flush response

Receiving
Station
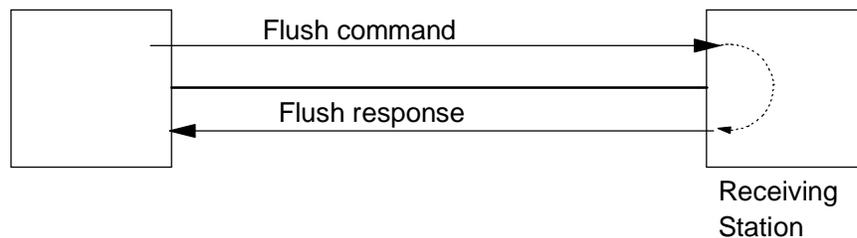
Flush Example

To move flow "C" to new link, flush is sent and subsequent "C" flow packets are held.
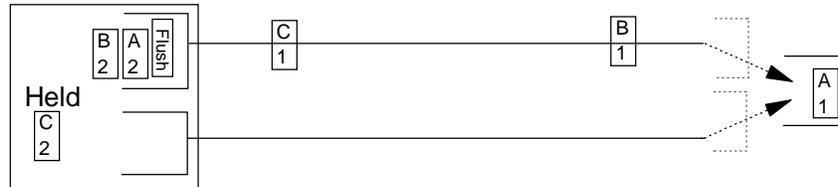
After flush response is received, C2 is queued for transmit

Scope of Flush Operation

- **Flush operation is from distributor to collector**
  - **Collector is responsible for managing its "output queue" to the MAC Client**
    - Queues are not explicitly addressed by the standard
      - The output queue is implicit in the operation of the collector
    - Collector must not respond until all data preceding the flush has been placed into its "output queue"
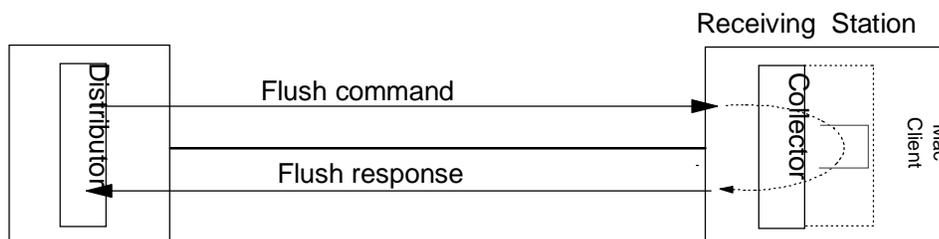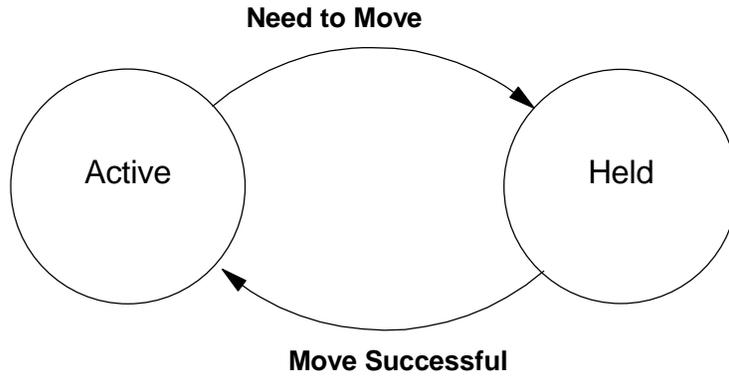
Flush Request/Response

- **If the Flush Response is returned on the same physical link that the Flush request was received on:**

    - **Simplifies implementations**
        - **Load balancing logic used for user flows is not involved in the determination of the return path for the flush response.**
        - **Flush response logic can be implemented at the collector**
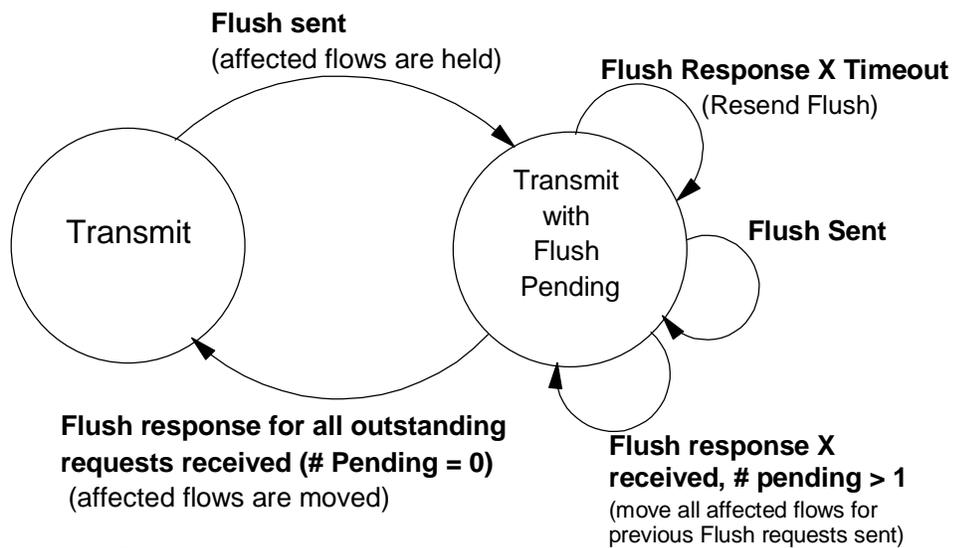   - **Positioned for incorporation into hardware**

Operational requirements for Flush

- Flush Commands go from distributor to collector
    - the "output queue" is an implicit part of collector operation

- Flush must be optional to send

- Response to the Flush command is Mandatory

- To accommodate various implementation choices for dealing with the flush scenarios, the Flush Command can be received at anytime (in any state).

- Flush is processed on a per link basis
    - 1-n flows can be flushed

State Transition Diagram
for
Individual Flows

Need to Move

Active → Held

Move Successful

Held → Active

Flush
Transmit State Diagram

Flush sent
(affected flows are held)

Flush Response X Timeout
(Resend Flush)

Transmit → Transmit with Flush Pending

Flush Sent

Flush response for all outstanding requests received (# Pending = 0)
(affected flows are moved)

Flush response X received, # pending > 1
(move all affected flows for previous Flush requests sent)

★ Flushing 1-n flows on a per link basis

Flush
Receive State Diagram

Flush received (Flush response sent after all preceding frames on this link have been placed in collector's "output queue")

Any State