On the Need for MAC Deferral During Fast Retrain

Related to 802.3az D3.1 Comments 83, 96, 97

David Chalupsky, Ilango Ganga
Intel Corp.

IEEE 802.3az Task Force
July 2010

Energy Efficient Ethernet

Supporters

- Sanjay Kasturia, Teranetics
- Hugh Barrass – Cisco
- Matt Brown, Brad Booth – Applied Micro
- Kamal Dalmia, Paul Langner – Aquantia

Energy Efficient Ethernet

Problem Statement

- Either IDLE or Local Fault may be generated by the PHY to the MAC during fast retrain.

- Neither option is desirable and has system-level impact.

Energy Efficient Ethernet

Option A – Sending IDLE to the MAC

In this case the MAC will be unaware that the link is not available.

- Pros: Avoids a Link Down indication to the host system.

- Cons: Allows 30msec of data from the MAC to be dropped silently.
- Violates the spirit of our 802.3az objective:
  - "No frames in transit shall be dropped or corrupted during the transition to and from the lower level of power consumption"
- Contrary to the Data Center Bridging efforts in 802.1 which seek to provide a "lossless Ethernet" fabric for datacenter applications.
- Makes 10GBASE-T a lossy protocol and will negate the DCB efforts.

Applications which use DCB expect a lossless infrastructure (i.e., FCoE).

Users of such applications will not accept 10GBASE-T with this behavior.

Energy Efficient Ethernet

Option B – Signal Local Fault to MAC during Fast Retrain

The MAC will interpret Local Fault as a link down condition.

Pros: Stops packet Tx from the MAC and minimizes data loss.

Cons:

- Creates a “link flap” condition as seen by higher layers.
- Typical response in a NIC application:
  - MAC sees Local Fault as link down
  - Controller interrupts driver
  - driver signals link down to operating system
  - Logs Link Down occurrence in system event log
  - Initiates fail over to redundant link
  - Possible workaround by filtering out link down indication for <30msec… but that delays fail over for legitimate link loss cases.
- Switching application will also see Local Fault as link loss with undesirable consequences, such as initiating failover.

Suggested Remedy

- Implement new message to RS which will indicate CARRIER_DETECT to MAC for Tx deferral.
    - <Link Unavailable>

- Utilize “MAC Transmit Deferral During Fast Retrain” as baseline solution.
  – brown_01_0710.pdf

Energy Efficient Ethernet

Buffering Considerations

<Link Unavailable> provides additional information to switch indicating temporary condition, allowing informed response.

- May choose to drop time-sensitive data (video/voice)
- May choose to buffer and pause no-drop traffic classes (i.e., FCoE)
- May choose only to pause best effort classes (TCP)

IDLE gives no indication of link condition.

- Data hopelessly lost.

Local Fault gives no indication that the condition is temporary.

- May prematurely begin discarding data, initiating fail over, etc.
- May choose to wait 30msec, in which case buffering impact is similar to having <Link Unavailable>
  - Delays failover for legitimate link loss case

Energy Efficient Ethernet