Joint meeting: 802.3az Energy Efficient Ethernet (EEE) Task Force and 802.1 Data Center Bridging (DCB) Task Group, Wednesday, March 19, 2008

Start time: 10:05 AM

Pat Thaler opened the meeting with an overview of the work being done in DCB

Task Group focus:
1- Priority based flow control
2- Congestion notification.

Meeting to discuss impact of EEE in data center traffic

Geoff: Seems like data center bridging uses Ethernet for connection. Is 802.1 doing anything to specify higher performance requirements?

Pat: Our project directed in FDX, Point to Point connection. We are only focused in Ethernet MAC.

Norm: AVB group having discussion with 802.11- trying to get .11 to be in the middle of network rather than the edge.

Mike: couple of ways we are considering to do EEE (see back up slides)

- Subset PHY technique
- LPI technique

First question:
- Buffer and burst issues and traffic misshaping. Looking for feedback on impact in data center.

Pat: The time duration of buffer and burst makes impact based on delay requirements.

Rob: It can be usec to msec range.

Hugh: The primary focus is for 100BASE-TX and 1000BASE-T, not 10GBASE-T.

Mike: How does asymmetric operation impact DCB?

Norm: Not sure.

Pat: For congestion notification, response to go back getting delayed could impact the throughput. Generally a couple of usec round trip delay for congestion notification is acceptable. If it is longer than that it could affect the system.

Norm: If there are multiple hops, then it will have impact.

Bob: Having traffic models would be beneficial. Also, EEE can be disabled. So, people can decide if they want more throughput or power saving. The edge connections benefit most on EEE, but the middle network data center less benefit from EEE.

Hugh: Good to know what is max latency for the return path needed.

Pat: for storage traffic, most traffic going in one direction than other. Return packets, e.g. ACKs need to be able to get back with minimal delay.

Mike: Market size for data centers?
Pat: see tutorials

Norm: AVB group doing bandwidth RSVP and need guarantee delivery. With bandwidth RSVP, need to include latency at each hop. Any buffer and burst will increase the size of buffer will increase latency and is accumulative at each hop.

Mario: Does AVB support 100BASE-TX and 10BASE-T?

Pat: AVB is speed-agnostic.

Pat: Uncompressed HD Video will require higher than 1G speed. The congestion management traffic is on 1G and 10G links and not at AVB. 200 usec delay for 10Gig round trip delay. For AVB, at 100M, need to reduce reduced packet size to make it work.

Mario: What is the latency requirement for AVB?

Rob: Fundamentally, is it better to send packets 10 time faster and then turn off the node than to run at lower speed?

Hugh: AVB excludes 10BASE-T and jumbo frame.

Pat: Reminder that will be meeting with a wider .1 group in July.

Norm: Congestion notification messages are short packets, concerned about latency with buffer and burst. It will impact latency, since few packets need to get to through at fast rate.

Norm: Everything depends on the flows. Every usec latency means the sessions need to stay open longer. No magic number. Every usec latency have impact.

Scott: What are the things you look at that would cause you to turn off the EEE mode?

Norm: In data centers, 2-5 usec will change from one technology to another. For cache coherent applications 2-5 is important. For other applications, can go down and not effect the network. So the management layer can shut down some paths that are not congested.

Geoff: what about link aggregation?

Norm: I was talking in general case. If they are link aggregated this does not apply.

Wael: Should one implementation for EEE apply to all applications? Do we need multiple knobs? This is a multi-dimensional problem. How much latency and power saving can be tolerated. Can we do buffering and save more power? Need help to determine how much EEE will affect data center.

Mitch: One of the possible knob is buffering and save more energy.

Hugh: The buffer needed is minimal. Delay a bit to have optimal power saving.

Caitlin: Concur with Wael. What you can do will drive the application. If big data center, it may be the EEE will not work and I will turn it off. EEE will certainly impact efforts. The implementation with multiple selections will be more useful than the one with single on-off mode.

Hugh: the Edge application is more important than the core. Need to know the latency distribution. One implementation is cyclical on-off, the other one is change speed. Short packet happen to after long frame, it will have impact to the overall.

Need to know:
Worst case latency
Average latency.

Manoj: The latency impact is when traffic starts.

Wael: End to end latency and hop to hop delay are important. Need to check with AVB group. Having multiple knobs is better than just ON-OFF.

Pat: If it is only on-off then may not be used at all.

Mario: Agreed. Need to consider the case where it is not congested network and utilization is low. IPG packet length distribution is important. Different time of day or months, the network may not be as used.

Wael: Why each case is exclusive? We can optimize power saving for very low traffic and less power saving as traffic BW increase.

Pat: Yes, we agree it is important to get data center traffic models. But data centers are not willing to share the info. Difficult to say this is typical or normal.
If congested, OK, don't use EEE? Traffic patterns may be bursty and symmetric.
Hard to decide which low speed node should be delayed since few packets may need to get through at fast rate due to priority.

Hugh: For a congested link, we can turn on fast enough.

Mitch: HPC network, Congestion management becomes important. Most HPC and data center willing to take penalty. If delay in usec then OK. For burstiness, Long bursts will be much longer.

Bob: Time of day or time of year? Defines how many AVB or nodes in data center will be used. There are case where a sync protocol, requires all nodes up and running….. Redundant links in data center can go in low power mode.

Manoj: Congestion mgmt does not need to run all the time in data center.

Mitch: Time of day or time of year? Power saving not useful by turning EEE ON/OFF. Traffic patterns should be used to turn EEE ON-OFF. This is more efficient to get most power saving. Operator will not turn ON-OFF.

Wael: If we can not implement this in a way that is used all the time rather than ON-OFF not very effective.

Pat: Traffic patterns can change sec to sec and msec to msec. So, should be able to implement a solution that can benefit from varying changes rather than ON-OFF based on time of day or year.

Mike: Regarding control mode, partially have some answers:

- Desire to have more control mode.
- More granularity knobs. Need to have power saving at all applications. IT can not determine when to turn ON or OFF.

Mike: What Granularity needed?

Mitch: Energy saving is good.
Norm: More Knobs are better.

Bob: Need to define protocol

Hugh: Better utilization of links, help with energy saving. Directing traffic to fewer links will impact energy saving.

Mitch Two modes to consider. Idle mode and active mode….

Pat: Need to know if the link is running at full speed or lower speed. For congestion notification, need to know link is running at lower speed or not.

Bob: In sleeping mode application, also need to know if link is sleep or not which effects congestion mgmt traffic.

Mitch: The timing of EEE could oscillate in usec …. Arrival rate and service rates are important consideration for going into EEE mode. Need to send info on rate of service, then we can use this info to apply to EEE control. Consider rate control info in congestion management to help with EEE application.

Geoff: Traffic models for MPG in Autos not realistic, but they had to use some models to start with. We just need to do some reasonable model to cover most cases.

Wael: Lets decide what to do next. Meeting in July what needs to take place between now and then?

Bob: Pushing for a joint meeting with .1 in July.

Mike: Based on questions come up here, hope people should come up with some actions. Hope some of our questions were answered.

Pat: Hope this was productive to make correct decision early on.

Mike: What is needed from EEE in July for the .1 meeting?

Pat: In July, do you plan to do a tutorial or just a meeting and discuss what is decided.

Mike, We provide update on what is decided.

Bob: By July, Hopefully will have a single solution. If not, better to have two full solutions. Should have some idea as how much buffering impact it will have.

Meeting adjourned at 12:00 PM