Technical Open Items for LPI

IEEE 802.3az, March 2008, Orlando, FL

Howard Frazier, Pat Thaler, Scott Powell, Wael Diab

Overview

- Objective: Attempt to summarize the technical issues raised by LPI which require further study
- Technical Issues split into 3 categories
  - PHY Related
  - Network Traffic Pattern Sensitivity
  - System Related

PHY Related Issues (1/2)

- Frequency drift, especially for 10GBASE-T
  - Frequency Lock and frequency drift may result in a long reacquisition.
    - direct impact on the transition time in and out of LPI
    - Frequency drift rate is not specified in 10GBASE-T or anywhere in 802.3
  - If you lose frequency lock then echo cancellation is lost
  - Above may be complicated by production and DVT tests that have temperature ramps
  - Please refer to grimwood_1_0308.pdf
- Scrambler in 100BASE-TX
  - Free running scramblers require 7 – 8 bytes (60 clean idle bits) to acquire sync
  - This may increase the wake-up time
  - In some cases, part of the wake up may thus require transmitting more idles. This would involve holding off the MAC until enough idles have gone through to other side
- Asymmetric noise related concerns
  - Cross talk during wake-up and refresh
  - When you do the refresh on the LPI path you may hammer the data active path

PHY Related Issues (2/2)

- BER may be degraded when leaving LPI in 10GBASE-T proposal
  - Assumption to use old parameters, start transmitting data even though BER is not right
    - On wake-up the proposal is to enter into PCS_Test
  - Slide 10 of parnaby_1_0108.pdf claims minimum 1msec which is based on minimum time for PCS_Test
  - >1 msec will then be the time to wake up not 10 usecs!
  - If PCS_Test fails then have to retrain which may require 100s of msecs. For example PMA_Fine_Adjust
    - Currently unbounded! Proposal in maintenance was 585 msecs average and the 650 msecs max for PMA_Fine_Adjust
  - This is similar to FastStart proposal
    - There is always the possibility that it doesn't converge
- Data integrity, and anything that degrades the BER up is not desirable
  - Not an acceptable solution to above or any other issues. Violates of our objectives

Traffic Pattern Sensitivity

- LPI could very well be useful on extremely low utilization links such as some 1G link to the desktop
- However, it is easy to generate traffic patterns that would render LPI significantly less effective, even under low traffic conditions
  - As an example, take a 1000 Byte Packet Stream at 8% of a 10G link
  - Packet size would be 1000 * 10e-10 * 8 = 0.8 usec
  - IPG would then be 0.8 usec * 92 / 8 = 9.2 usec
  - PHY could never get into LPI as transition time > IPG
  - Savings harder to predict or guarantee
- Above example could be further complicated in a real life network
  - 8% could be made up of multiple streams that are non-phase correlated, resulting in further less predictable (at least for a mid-point switch) chopping of the IPG
  - If the packet size is reduced for the same utilization then sensitivity is increased
- Similar scenarios can occur for 1G links as well
- Sensitivity could be mitigated/alleviated by “Buffer and Burst”. “Buffer and Burst” gives rise to separate concerns, see system issues

System Related Issues (1/2)

- Buffer (batch) and burst
  - "Traffic misshaping" concerns: Current direction within industry is to smooth out traffic and not to batch and burst
  - Impact on congestion
    - If multiple switches do this, may introduce congestion into an otherwise "quiet", un-congested and happy network
  - Impact on QoS
    - What happens if packets have different QoS profiles?
    - Is it acceptable to delay the packets blindly and if not, will the feature be disabled?
- Negotiation of "wake-up" time
  - Direct impact on the usefulness of buffering. Expensive system with high buffering to accommodate variable wake-up times would be limited by other cheap side
  - Control policies are affected by wake-up time. If there is a high variance on wake-up time then there would be an impact on the effectiveness of the policy or the policy would have to accommodate a wide range, which may be impractical (10usec to 1msec)
  - Testing would be complicated by this
  - Prefer to be bounded or limited variance

System Related Issues (2/2)

- Jitter increase and variation due to buffering
  - Increased delay variation due to batching and bursting that is policy and traffic dependent
  - May not be acceptable for switches that are in the middle of a network. Switches differentiated by time for a packet to go into fabric and out the other port. Buffering changes that
- Buffer flush and timer complexity
  - Idea of a lone or “orphan” packet sitting in the buffer and no other traffic
  - Usually done at higher layers
- Solution as presented appears to be optimized for desktop links not AVB enabled, Data Center, or switch to switch (midpoint) links
  - Traffic patterns for other markets / applications like those stated above are very different from low utilization bursty links, router and/or backbone links
- Any EEE solution should be simulated with an interactive command-response type protocol that involves response times for example iSCSI, FCoE etc.

Additional Work Being Considered

- Traffic patterns for non-edge devices and DC edge devices
  - Welcome help in gathering such data
  - Goal is to have a comprehensive set of models to evaluate all proposals against
- Up-shift transition during max length packet
- What level of buffering is acceptable for latency sensitive applications, CBR applications and/or both?
- Anything else missing from this list?

Concluding Remarks

- Not opposed to LPI
- Indeed, there are scenarios / markets / applications such as 1G desktop links where LPI is very attractive. However, in its current form, LPI does not appear to be optimal for AVB enabled, Data Center, or switch to switch (midpoint) links
  - Concern is that users would “turn this feature off”
- Technical issues raised reflect broader applicability of EEE and impact throughout the system
  - Looking forward to jointly address these issues
- Welcome feedback on Subset PHY and/or other goals / desires / concerns with EEE