Energy Efficient Ethernet & A/V Bridging

Bob Grow

Chair, IEEE 802.3

Intel Corporation

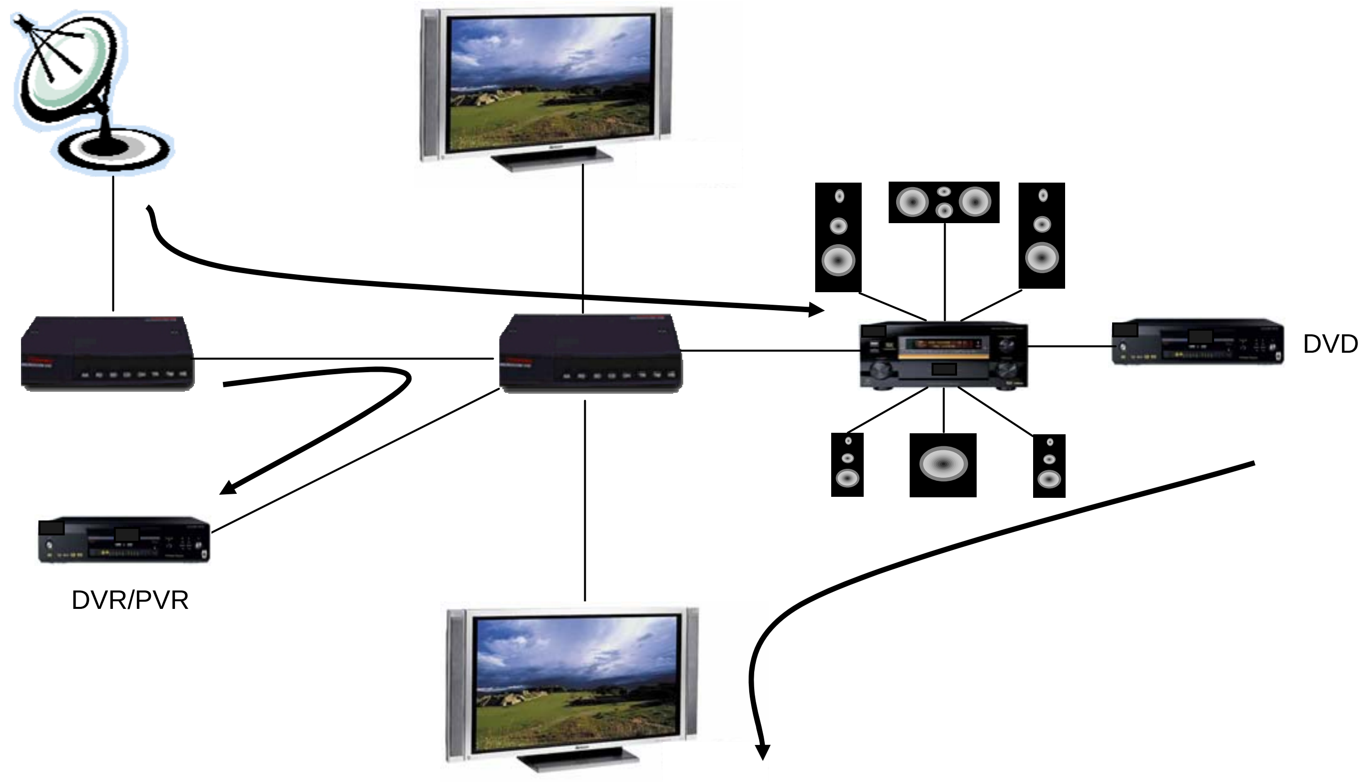
This presentation steals liberally from the July 2007 EEE plenary material (Barrass, Bennett, Diab, Law, Nordman)

IEEE 802.1 Interim
Stockholm Sweden
September 2007

What is Energy Efficient Ethernet?

- **A method to reduce energy use by an Ethernet interface rapidly changing to a lower link speed during periods of low link utilization**
- **Focus applications for energy reduction**
  - **Datacenter – Energy Star requirements go into effect in 2009**
  - **Enterprise horizontal networks – System reductions currently served with wake-on-LAN**
  - **Home – Enable designs that lower average energy consumption for always on networks**
  - **Emerging applications – Especially residential**
- **EEE will not be defining the control policy for link speed change, only the capability and hooks for control policy use**

Application - EAV home network



1. Listening to satellite radio on EAV receiver, link between receiver and switch operating at 10 Mbps
2. Start playing DVD on a screen in another room
3. Link between receiver and switch must transition from 10 Mbps to 100 or 1000 Mbps
4. Transition time must be less than 10 ms to avoid audible disruption
5. DVR/PVR set to record "Survivor" from satellite receiver at 8:00 pm on Thursday
6. Link between satellite receiver and AVB switch must transition from 10 Mbps to 100 or 1000 Mbps
7. Transition time must be less than 10 ms to avoid audible disruption

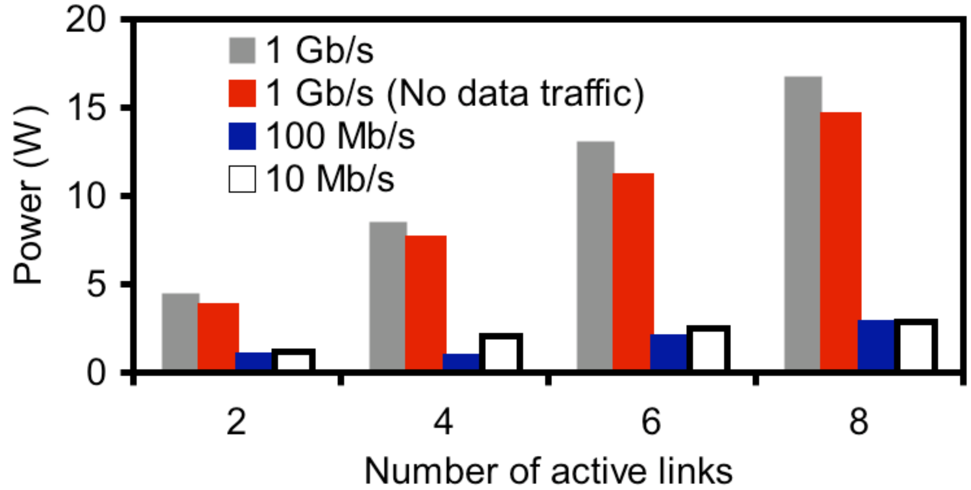
Link power

Results from (rough) measurements

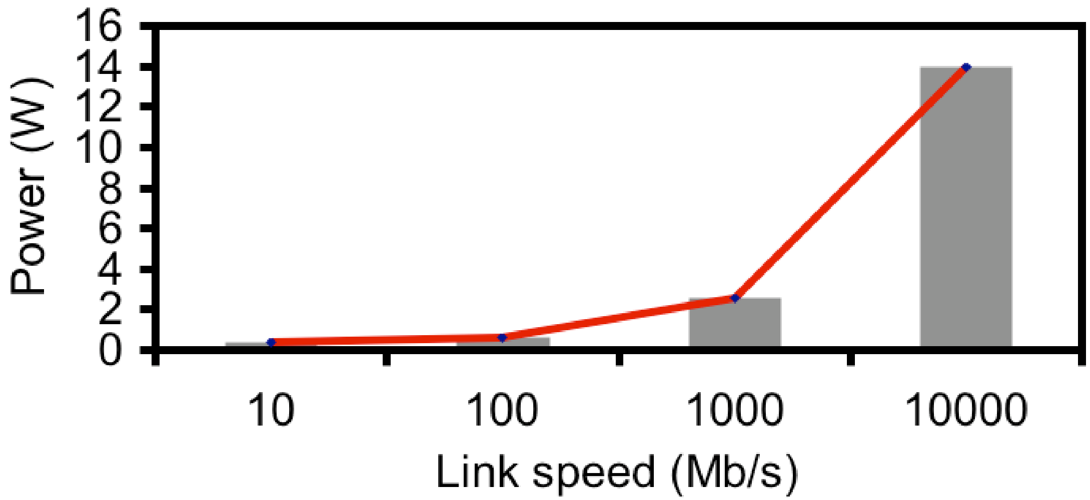
— all incremental AC power

— measuring 1st order

- **Typical switch with 24 ports 10/100/1000 Mb/s**



- **Various computer NICs averaged**



Potential Savings from EEE

Assume 100% adoption (U.S. Only), PHY energy savings only

- **Residential**
  - **PCs, network equipment, other**
  - **1.73 to 2.60 TWh/year**
  - **$139 to $208 million/year**
- **Commercial (Office)**
  - **PCs, switches, printers, etc.**
  - **1.47 to 2.21 TWh/year**
  - **$118 to $177 million/year**
- **Data Centers**
  - **Servers, storage, switches, routers, etc.**
  - **0.53 to 1.05 TWh/year**
  - **$42 to $84 million/year**

These figures do **not** include savings from cooling, power infrastructure, other system components, etc.

Total: $298 to $469 million/year

Study Group Objectives

- **Define a mechanism to reduce power consumption during periods of low link utilization for the following PHYs**
  - **100BASE-TX (Full Duplex)**
  - **1000BASE-T (Full Duplex)**
  - **10GBASE-T**
  - **10GBASE-KR**
  - **10GBASE-KX4**
- **Define a protocol to coordinate transitions to or from a lower level of power consumption**
- **The link status should not change as a result of the transition**
- **No frames in transit shall be dropped or corrupted during the transition to and from the lower level of power consumption**
- **The transition time to and from the lower level of power consumption should be transparent to upper layer protocols and applications**
- **Define a 10 megabit PHY with a reduced transmit amplitude requirement such that it shall be fully interoperable with legacy 10BASE-T PHYs over 100 m of Class D (Category 5) or better cabling to enable reduced power implementations**
- **Any new twisted-pair and/or backplane PHY for EEE shall include legacy compatible auto negotiation**

Reducing the link rate

- **Can (and does) save energy**
- **Some NICs drop link rate when a laptop is battery powered**
  - **Or, when a PC goes into sleep state**
  - **Turns-off PHY if no signal on link**
- **Match the link rate to utilization**
  - **High utilization = high link rate**
  - **Low utilization = low link rate**
- **Currently implemented using auto-negotiation**
  - **Set the Technology ability bits/message codes and then reset the link**
  - **Takes about 1000 milliseconds (a looooooong time)**
- **EEE targets a fast speed change**

Goals for EEE Transitions

- **Be safe: do no harm**
  - **Base results on WORKING systems**
  - **No change to operational mode of existing PHYs**
- **Be lazy: don't invent unnecessary things**
  - **Transition would minimally impact existing specifications**
    - **Reuse of existing 802.3an PHY control as much as possible**
- **Be quick: get PHY transition times down**
  - **Need for transitions of <10msec, pref ~1msec**
  - **Need to minimize retraining time**

Matching Link Rate to Utilization

- Rapid PHY Selection (RPS) includes
  - A PHY selection mechanism
  - A control protocol

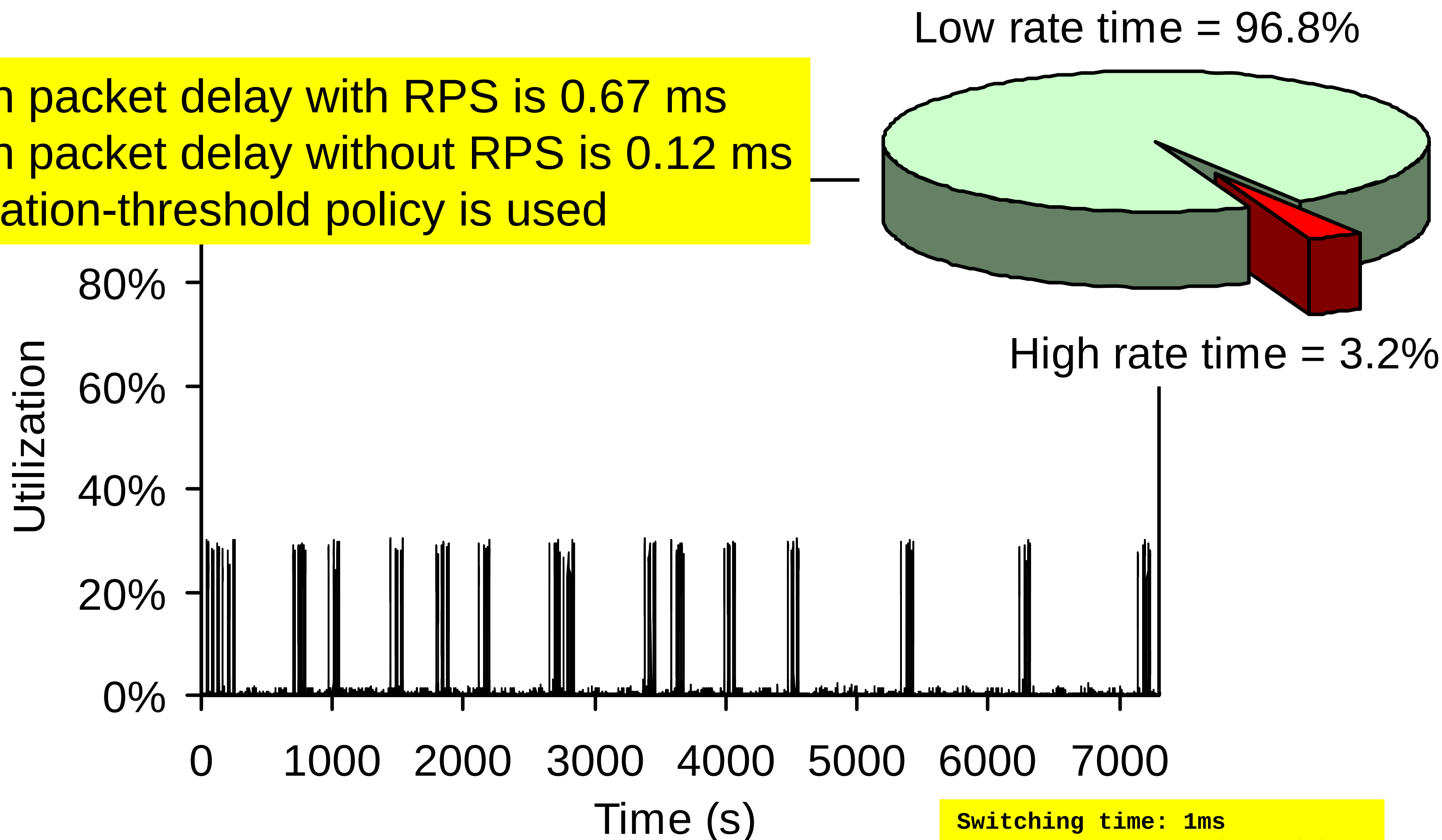
- PHY-level challenges
  - How fast to handshake?
  - How to re-synchronize for 1 Gb/s
  - How to re-synchronize for 10 Gb/s?

- Speed control policy is an area in which we could use some help

RPS – a picture tells the story

- **Snapshot of a typical Ethernet link with simulated RPS**

Mean packet delay with RPS is 0.67 ms
Mean packet delay without RPS is 0.12 ms
Utilization-threshold policy is used




```
Switching time: 1ms
Data rates: 100 and 10 Mb/s
Low/High thresh: 0KB and 32KB
```

Alternative to RPS

- Some people are concerned about the impact of transition time on applications
- An initial study on feasibility of 1 ms transition from lower speed to 10GBASE-T suggested 20 ms was feasible, 1 ms was not
- More concerns raised regarding impact on real-time applications such as Audio Video Bridging (AVB)
  - Transition time needs to be at most 1 ms
  - The problem is initial experiments suggested 20 ms or more for higher speed Ethernet PHYs
- What are AVB requirements for EEE?

AVB link utilization

- **Existing evidence of low utilization (desktop users)**
  - **LAN link utilization is generally in range 1 to 5% [1, 2]**
  - **Utilization for "busiest" user in USF was 4% of 100 Mb/s**
- **AVB links will also be low average utilization**
  - **Time of day usage patterns**
  - **Typical Ethernet overcapacity**
- **How do we expect utilization differ in AVB networks?**

[1] A. Odlyzko, "Data Networks are Lightly Utilized, and Will Stay That Way", *Review of Network Economics*, Vol. 2, No. 3, pp. 210-237, September 2003.

[2] R. Pang, M. Allman, M. Bennett, J. Lee, V. Paxson, and B. Tierney, "A First Look at Modern Enterprise Traffic," *Proceedings of IMC 2005*, October 2005

10BASE-T

- **10BASE-T was standardized in 1990**
  - **Medium: Category 3 Unshielded Twisted Pair (UTP)**
  - **Estimated less than 3% of installed base of cable is Category 3 / DIW**
- **Modify the specification to allow for transmission over Category 5 cabling or better**
  - **10BASE-T is the only popular Ethernet PHY not continuously clocked.**
  - **Chip vendors can eliminate the power rail currently only used for 10BASE-T**
- **Should it be an AVB standby speed?**

A wider view

- EEE study group has discussed saving power in the PHY
  - RPS – or similar mechanisms
- But whole system power will follow
  - Power savings vs PHY speed > expected PHY power
  - Even existing systems are saving more than PHY power
- Examine current and potential system power savings
  - "Reduction of power during low link-utilization"
  - Where will this benefit from standards-based control?

EEE implications to AVB

- **Link unavailable during speed change**
  - **EEE will produce short term (1ms?) link unavailability**
  - **Frames will be delayed during link unavailability unless discarded**
- **Link speed will produce latency variation**
  - **Transmission time of frame**
  - **Latency of component data paths**
- **Speed change affects available bandwidth**
  - **Obvious impact on reservation protocol**
- **Power state may need to be an AVB consideration**
  - **Quick start as devices become sentient**
  - **Grandmaster selection**
  - **Can EEE be enabled on any but edge links**

More information

- **July plenary tutorial**
  - **http://www.ieee802.org/802_tutorials/july07/IEEE-tutorial-energy-efficient-ethernet.pdf**
- **EEESG archive**
  - **http://www.ieee802.org/3/eee_study/index.html**