

Energy-Efficient Ethernet for 40G/ 100G Next Generation Optics

IEEE 40G and 100G Next Generation Optics Study Group

Michael J. Bennett
Lawrence Berkeley National Laboratory

Wael William Diab
Broadcom Corporation

Contributors and Supporters (partial list)

- Hugh Barrass, Cisco

Topics

- Brief Recap of EEE
- Motivation for EEE in this project
- How this could apply to 40G/100G Next Generation Optical Ethernet

EEE Recap

- Energy Efficient Ethernet (EEE) is a method to reduce energy used by an Ethernet device during periods of low link utilization
- Specified in IEEE Std 802.3az-2010™
- The premise for EEE is that Ethernet links have idle time and thus opportunity to save energy
- Specified for copper interfaces
 - “BASE-T’ s’
 - Backplane (except 40G)
- The method is called Low Power Idle (LPI)
 - See [bennett_01_0311.pdf](#) for more details

Motivation for EEE in this project

- Reduced power consumption contributes to market potential
 - Lane reduction lowers component count, which in turn, lowers power consumption
 - Lower power consumption enables higher port density, which in turn, lowers cost to the end user
 - EEE can enhance energy-savings opportunity
 - Savings is proportional to utilization

Source: http://www.ieee802.org/3/100GNGOPTX/public/jul11/CFI_01_0711.pdf#page=17

Source: http://www.ieee802.org/3/100GNGOPTX/public/jul11/CFI_01_0711.pdf#page=18

Motivation for EEE in this project

- EEE could help make the datacenter more energy proportional to load¹.
- End users are asking developers to “make better energy proportionality a primary design objective” for future systems¹.
- Savings for the IEEE 802.3az PHY alone should be around 90% and energy reduced by up to 70% for the NIC when in LPI mode².
 - much greater savings possible in systems using LLDP
 - See dove_02_05_08.pdf (slide 5)

1. L. Barroso and U. Hölzle, The Case for Energy-Proportional Computing. Computer, 40(12):33-37, December 2007
2. P. Reviriego, K. Christensen, J. Rabanillo, and J. A. Maestro, 'An Initial Evaluation of Energy Efficient Ethernet' in IEEE communications letters, VOL. 15, NO. 5, May 2011

Motivation for EEE in this project

- Data center operators are very much interested in using power efficiently as energy-use impacts operational expense
 - E.g. Google spent ~\$200M on energy in 2010
 - Note that Google's data centers are roughly 50% more efficient than others
- Data Center Operators want energy-proportional equipment
- Larger data centers use optical links
 - “Likely a lot of value in figuring out EEE for optical links”

Source: http://www.ethernetalliance.org/wp-content/uploads/2012/02/EATEF_Panel-3_Power_12_0216.pdf
slides 51,52,56

Motivation for EEE in this project

- Energy Efficiency is a priority for regulators
 - EU CoC on Energy Consumption of Data Centers
 - Energy Star specs for Small Network Equipment
 - Large Network Equipment coming
 - Policy will encourage technologies like EEE
 - Can support that by including EEE in the specification
- Therefore EEE is a "must" for a new specification

Source: http://www.itu.int/dms_pub/itu-t/oth/09/05/T09050000010004PDFE.pdf

Source: http://www.energystar.gov/index.cfm?c=new_specs.small_network equip

Motivation for EEE in this project

- EEE should be included *at the beginning of projects*
 - Very difficult and time consuming task to retrofit EEE into completed specifications
 - Run the risk of breaking things
 - Much more efficient to consider EEE in the initial specification

How could this apply to 40G/100G NG Optics?

- Lowest hanging fruit for 40G and 100G NG Optics
 - Use LPI codewords for signaling – no PMD power-down
 - Bulk of the work is being done in P802.3bj
- EEE is showing up in switches and will likely be a feature in most, if not all Ethernet switches by the time this project is finished
 - Including EEE in this project enhances market potential
- Is there interest in working on EEE?

Thank You!

backup

Motivation for EEE in this project

- Energy cost is still a significant operational expense in data centers

Company	Servers	Electricity	Cost
eBay	16K	$\sim 0.6 \times 10^5$ MWh	$\sim \$3.7\text{M}$
Akamai	40K	$\sim 1.7 \times 10^5$ MWh	$\sim \$10\text{M}$
Rackspace	50K	$\sim 2 \times 10^5$ MWh	$\sim \$12\text{M}$
Microsoft	>200K	$> 6 \times 10^5$ MWh	$> \$36\text{M}$
Google	>500K	$> 6.3 \times 10^5$ MWh	$> \$38\text{M}$
USA (2006)	10.9M	610×10^5 MWh	$\$4.5\text{B}$
MIT campus		2.7×10^5 MWh	$\$62\text{M}$

Source: Cutting the Electric Bill for Internet-Scale Systems, Qureshi et. al, SIGCOMM '09 Proceedings of the ACM
SIGCOMM 2009 conference on Data communication, ISBN: 978-1-60558-594-9

Motivation for EEE in this project

- Even in high transaction-rate networks, utilization is not 100% 24 hours/day, 365 days/year = opportunity to save energy

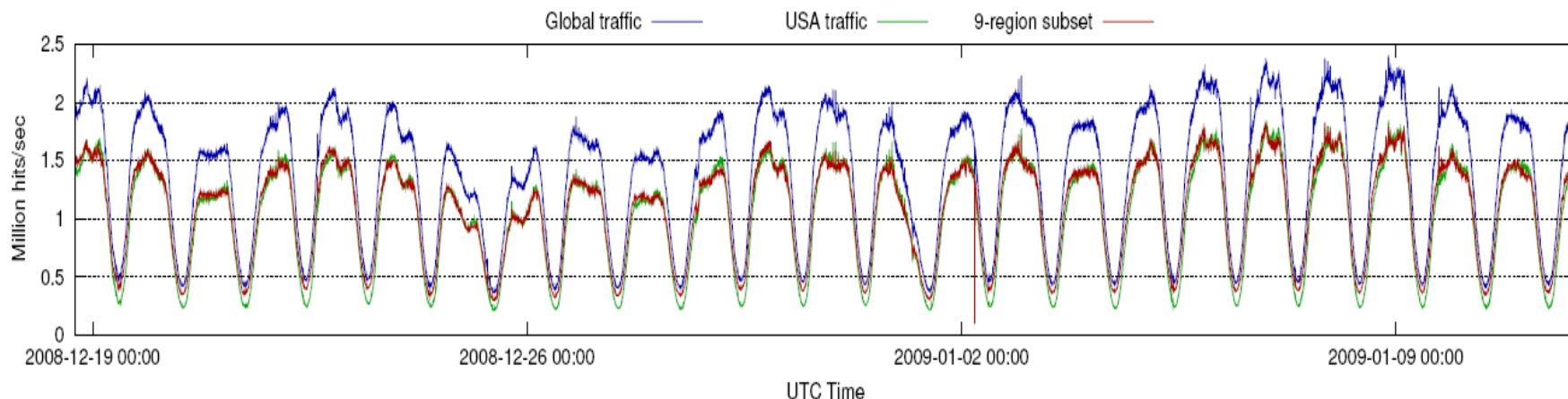


Figure 14: Traffic in the Akamai data set. We see a peak hit rate of over 2 million hits per second. Of this, about 1.25 million hits come from the US. The traffic in this data set comes from roughly half of the servers Akamai runs. In comparison, in total, Akamai sees around 275 billion hits/day.

Source: Cutting the Electric Bill for Internet-Scale Systems, Qureshi et. al, SIGCOMM '09 Proceedings of the ACM
SIGCOMM 2009 conference on Data communication, ISBN: 978-1-60558-594-9

Open Questions

- What about service providers selectively powering off elements of data centers?
 - Does it make sense to implement LPI if data center operators are selectively powering off elements, e.g. containers, racks, servers, etc?
 - The idea has been suggested for storage systems
 - This appears to be a hot research topic, but it is difficult to find examples of application in production use.
 - The answer depends on the future development of energy management applications for data centers

Source: http://www.cs.cornell.edu/projects/quicksilver/public_pdfs/IEEE%20NCA%202011.pdf

Open Questions

- Does EEE for 40G/100G Next Generation Optics mean powering off the PMDs?
 - Not necessarily
 - Could allow LPI signaling to pass the lower utilization state to the link partner
 - Use it to reduce system power consumption
 - Can take advantage of the Ethernet switch port savings for a wider variety of port types