Energy-Efficient Ethernet for 100G Backplane and Copper

IEEE 802.3bj Task Force

Michael J. Bennett
Lawrence Berkeley National Laboratory

Contributors

Hugh Barrass, Cisco
Luiz Andres Barroso, Google
Brad Booth, Dell
Matt Brown, Applied Micro
Ken Christensen, University of South Florida
Robert Coenen, Reflex Photonics
Wael Diab, Broadcom
Dan Dove, HP
Bob Felderman, Google
Alan Flatman, LAN Technologies
Ilango Ganga, Intel
Bob Grow, Intel
Mark Gustlin, Cisco
Rob Hays, Intel
Steven Lanzisera, LBNL
David Law, HP
Jeff Lynch, IBM
Inder Monga, ESnet
Shimon Muller, Oracle
Bruce Nordman, LBNL
David Ofelt, Juniper
Geoff Thompson, GraCaSI S.A
Pedro Reviriego Vasallo, Nebrija Univ.

Topics

- Introduction
- Brief Recap of EEE
- Motivation for EEE in this project

Introduction

- Energy Efficient Ethernet (EEE) is a return to old principles
  - Original coax Ethernet was based on a basic EEE principle, i.e. no network power was exerted between packets
    - this changed with the active idle on FOIRL
- EEE principle #2 is that higher speeds during data transmission use less energy
    - Smaller bit times
    - Good to go to higher speeds from an energy perspective
- EEE drives designs back to these principles as much as possible.
- I'm asking that new projects embrace these concepts from the outset

EEE Recap

- 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 up to 10G
- The method is called Low Power Idle (LPI)
  - See bennett_01_0311.pdf for more details

Motivation for EEE in this project

- Backplane and copper systems as defined in this project will be used primarily in data centers
  - Number of blade servers forecast to increase by 10% CAGR from 2011 to 2015 [3]
  - First generation 100G servers forecast to hit the market by 2016 [4]
    - Note – typically, first generation technology uses more energy than subsequent generations

3. http://www.ieee802.org/3/100GCU/public/nov10/CFI_01_1110.pdf#12
4. http://www.ieee802.org/3/100GCU/public/nov10/CFI_01_1110.pdf#11

Motivation for EEE in this project

- Energy cost is still a significant operational expense in data centers [1]

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

1. 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. Estimated annual electricity costs for large companies (servers and infrastructure) @ $60/MWh (6 cents / KWh)

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[1]

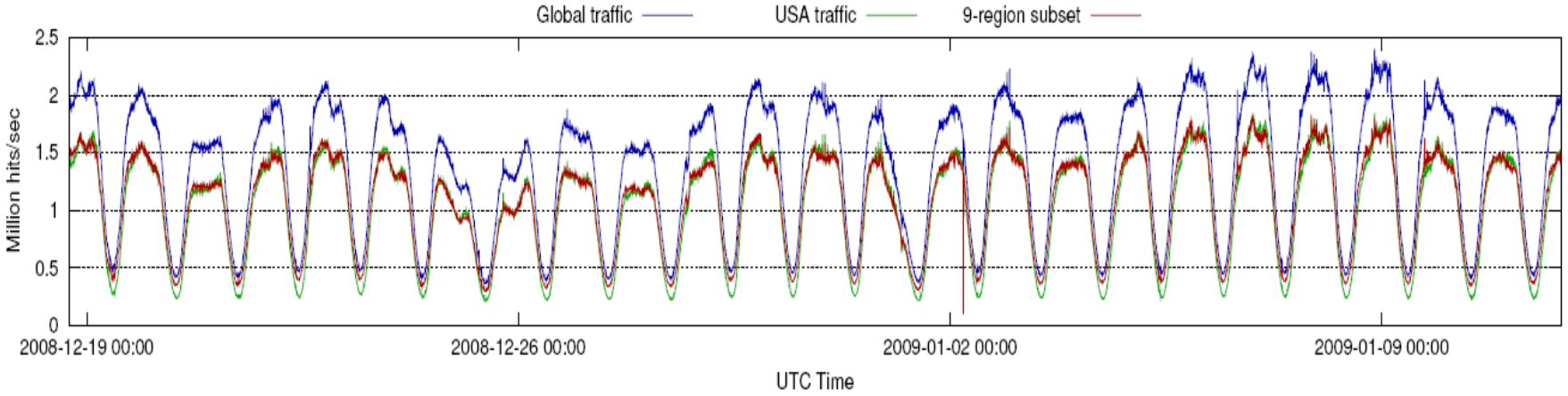

Figure 14: Traffic in the Akamai data set. We see a peak hit rate of over 2 million hits per second. Of this, about 1.25 million hits come from the US. The traffic in this data set comes from roughly half of the servers Akamai runs. In comparison, in total, Akamai sees around 275 billion hits/day.

1. 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

- Another example of an energy saving opportunity

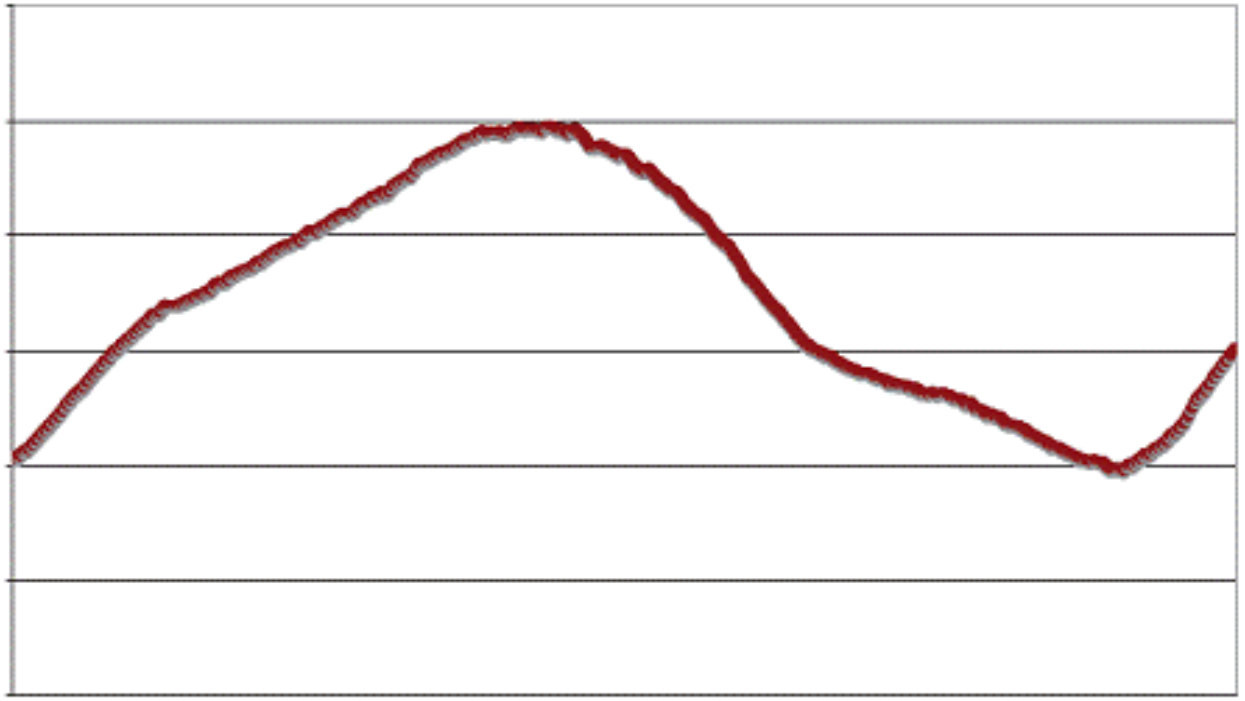

FIGURE 2.2: Example of daily traffic fluctuation for a search service in one datacenter; *x*-axis is a 24-h period and the *y*-axis is traffic measured in queries per second.

1. http://research.google.com/pubs/pub35290.html

Motivation for EEE in this project

- Energy Efficiency is a priority for regulators
  - EU Code of Conduct on Energy Consumption of Data Centers[1]
  - Energy Star specs for Small Network Equipment[2]
    - Large Network Equipment coming
  - Policy will encourage technologies like EEE
  - can support that by including EEE in design
- Therefore EEE is a "must" for a new design

1. http://www.itu.int/dms_pub/itu-t/oth/09/05/T09050000010004PDFE.pdf
2. http://www.energystar.gov/index.cfm?c=new_specs.small_network_equip

Motivation for EEE in this project

- EEE could help make the datacenter more energy proportional to load[1].
- End users are asking developers to “make better energy proportionality a primary design objective” for future systems[1].
- 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 mode[2].
  - 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

- EEE should be included *at the beginning of projects*
  - Very difficult and time consuming task to retrofit EEE into completed specifications
    - we are only looking at 100G not 40G here
  - Run the risk of breaking things
  - Much more efficient to consider EEE in the initial design
- Inclusion of EEE should be considered similar to auto-negotiation
  - You need it, just do it

Things to consider

- Latency
  - 10GBASE-KR LPI has a wake time of 11.25 usec
  - 100G BP PHY?
  - 100G Cu PHY?
  - Feedback I received stated “to be broadly useful the power-on delay would need to be in the low usec range” (single digit)
  - Merits and considerations of reusing LPI
  - leverage the work done in 802.3az
- Merits and considerations of alternative technologies?

Thank You!