Improving the Energy Efficiency of Networks: A Focus on Ethernet and End Devices

Ken Christensen
Computer Science and Engineering
University of South Florida
Tampa, FL 33620
christen@cse.usf.edu
(813) 974-4761

Bruce Nordman
Energy Analysis
Lawrence Berkeley National Laboratory
Berkeley, CA 94720
BNordman@LBL.gov
(510) 486-7089

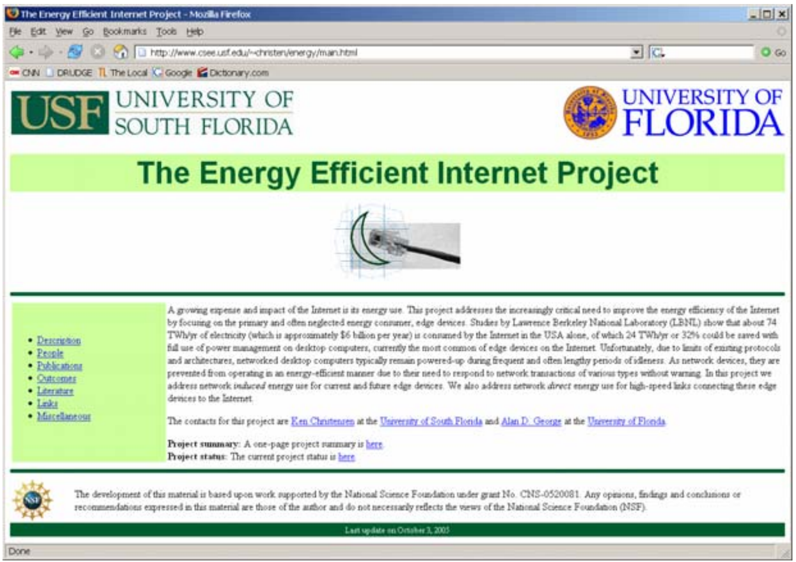
This material is based upon work funded by the National Science Foundation under grant CNS-0520081 for Christensen

Support from the EPA Energy Star Program and the California Energy Commission Public Interest Energy Research Program

Acknowledgments

- **We thank Flavio Bonomi for inviting us**

Our project…



http://www.csee.usf.edu/~christen/energy/main.html

Agenda

- Energy use by IT equipment
- Regulatory/policy directions
- Reducing *direct* network energy use
- Reducing *induced* network energy use
- Summary and future directions

Agenda

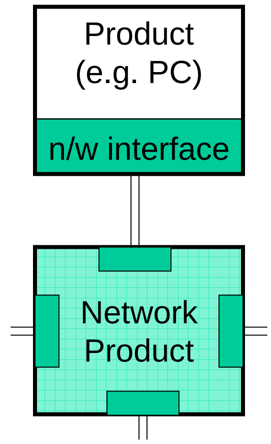
- **Energy use by IT equipment**
- **Regulatory/policy directions**
- **Reducing *direct* network energy use**
- **Reducing *induced* network energy use**
- **Summary and future directions**

Electronics energy context

- **1 TWh/year = $80 million/year**
  - (U.S. average is $0.08 / kWh)
- **Annual Consumption** (circa 2006)
  - All electronics: ~ 200 TWh — $16 billion
  - Networked electronics: ~ 100 TWh — $8 billion
  - Network equipment & NICs: ~ 25 TWh — $2 billion
  - Data center energy use: ~ 35 TWh — $3 billion
- **Caveats**
  - Figures not well known — existing estimates are old
  - U.S. only — multiply severalfold for rest of world
  - Does not include cooling or power infrastructure

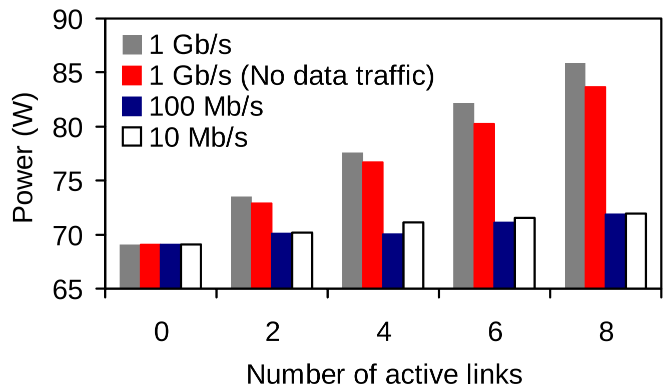
Network energy context

- **Network Direct Energy Use**
  - NICs
  - Network Products
    - Switches, Routers, etc.
- **Network Induced Energy Use**
  - Increment for higher power state of devices needed to maintain network connectivity (usually On instead of Sleep or Off)
  - Common causes:
    - Can't maintain needed connectivity
    - Too cumbersome to set up or use
  - Key products: PCs and Set-top boxes



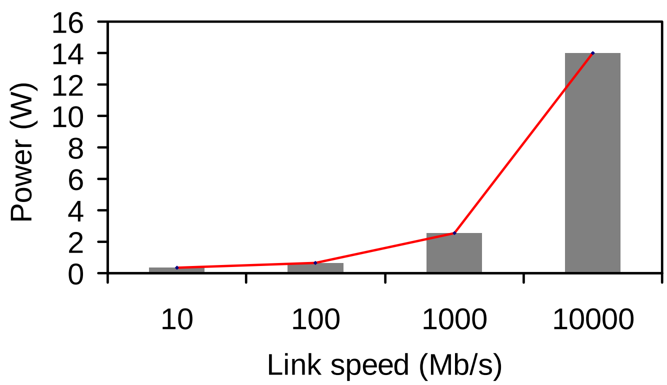
Measured power use of switch ports

- **Cisco Catalyst 2970 with 24 10/100/1000 Mb/s ports**
  - Measurements taken at the wall



Measured power use of NICs

- **Various NICs averaged**
  - Measurements taken at the wall



Trends in network energy consumption

- **More connections** ⬆
  - Consumer electronics increasingly (digitally) networked; IPTV
- **Longer on-times** ⬆
  - "Always available"
- **Higher Speeds** ⬆
  - Video
- **More network technologies** ⬆
  - Products with multiple connectivity
- **More energy-efficient** ⬇
  - Less power for fixed speed

Bottom line: Network energy burden is rising

Trends continued

- **"All electronics are lightly utilized"**
  - >2/3 of PC energy use when no one present
  - Typical commercial server utilization: ~15 to 20%
  - Typical network link usage:
    ~3 to 5% [Odlyzko]; 1 to 5% [Pang]
    - Edge links even lower
- **Energy not traditionally a major design criterion**
  - At least not *average* energy use

Agenda

- **Energy use by IT equipment**
- **Regulatory/policy directions**
- **Reducing *direct* network energy use**
- **Reducing *induced* network energy use**
- **Summary and future directions**

Regulatory / policy / industry trends

- **IT industry**
  - Increasing pressure to address energy/env. impacts
  - Concern over data center energy/power problems
- **Energy and Environment**
  - Electronics and Misc. a growing % of all electricity
  - Climate impact a rising concern
- **Energy Policy**
  - Energy Star: networks a rapidly growing topic
  - Diverse international standards/labeling

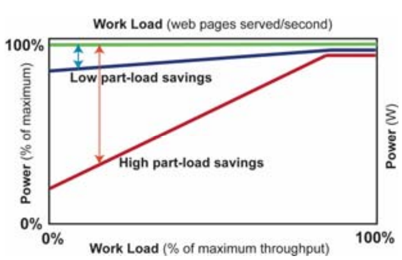
109TH CONGRESS 2d Session | HOUSE OF REPRESENTATIVES | REPORT 109–538

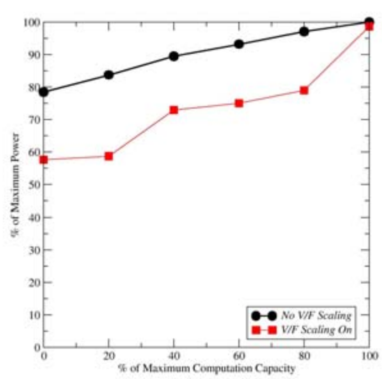
ENERGY EFFICIENT COMPUTER SERVERS

the green grid™

Server power metrics

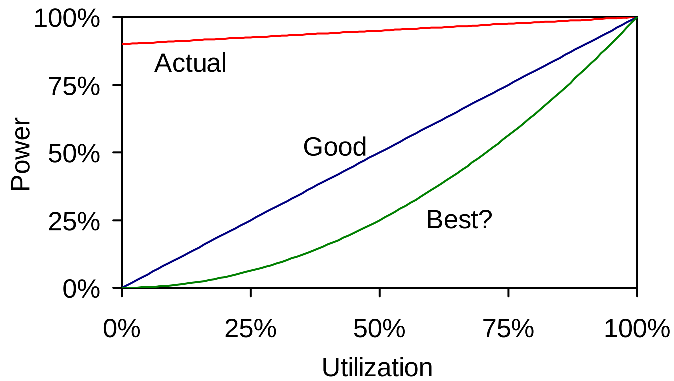
- **Server scaling of power to computational load**
  - Methods
  - Data
  - Products





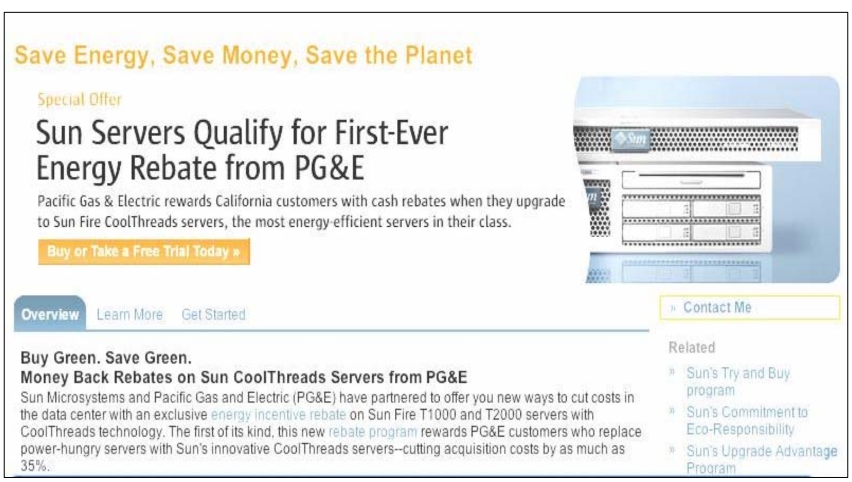
Actual versus ideal power use

- **Power versus utilization for network equipment**
  - Can V/F scaling and other technologies can achieve “Good” or better?



Energy industry responds

- **Rebates for buying efficient servers**



Link rate changes in ADSL2

- **ADSL2 is a last mile "to the home" technology**
  - 30 million DSL subscribers worldwide
- **ADSL2 is G.992.3, G.922.4, and G.992.5 from ITU**
  - Standardized in 2002
- **ADSL2 supports power management capabilities**
  - Link states L0 = full link data rate
  - Link state L2 = reduced link data rate
  - Link state L3 = link is off

Energy Star requirements

Computer specification, being finalized NOW!

- **Tier I (2007):**

  Computers shall reduce the speed of any active 1 Gb/s Ethernet network links when transitioning to Sleep or Standby.

- **Tier II (2009):**

  ENERGY STAR qualified computers must retain full network connectivity while in Sleep mode, according to a platform-independent industry standard.

  All computers shall reduce their network link speeds during times of low data traffic levels in accordance with any industry standards that provide for quick transitions among link rates.

Agenda

- **Energy use by IT equipment**
- **Regulatory/policy directions**
- **Reducing *direct* network energy use**
- **Reducing *induced* network energy use**
- **Summary and future directions**

Links are lightly utilized

- **Desktop-to-switch links**
  - Are mostly idle
    - Lots of very low bandwidth "chatter"
  - High bandwidth needed for bursts
    - Bursts may be seconds to minutes (to hours!) apart
- **Server links are also often not fully utilized**
- **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

[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

Links are lightly utilized <u>continued</u>

- **Snapshot of a typical 100 Mb/s Ethernet link**
  - Shows time versus utilization (trace from Singh at PSU)

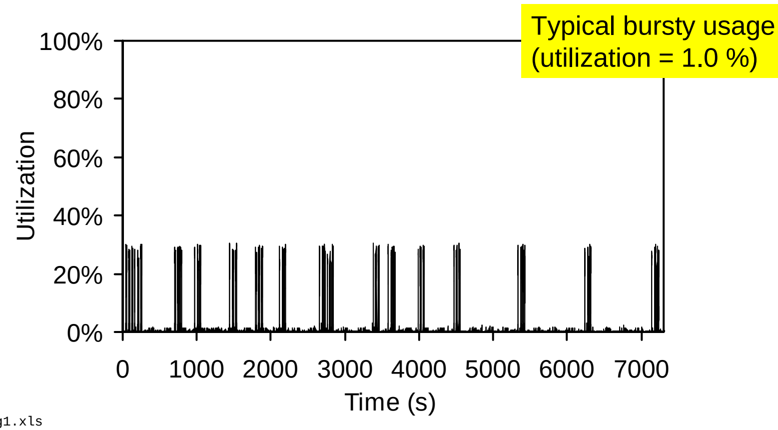


fig1.xls

Reducing the link rate

- **Match the link rate to utilization**
  - High utilization = high link rate
  - Low utilization = low link rate

- **Can (and does) save energy**
  - Intel NICs drop link rate when a notebook is battery powered or when a PC goes into a sleep state
    - Turn-off completely if no signal on link
- **Currently implemented using auto-negotiation**
  - Set the SPEED bits and then reset the link
  - Takes at least 250 milliseconds (a looooong time)

* From Intel 82641PI product information web site (2005)

Need for fast link transitions

- **Can extend the benefits of link data rate reduction**
  - By making the data rate transition faster
- **Need a faster *mechanism* than auto-negotiation**
- **Need a *control policy* to determine when to**
  - Transition from low to high data rate
  - Transition from high to low data rate

Adaptive Link Rate

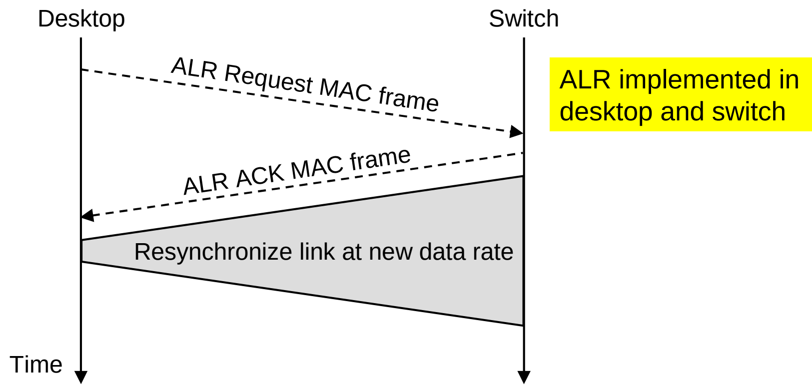
Open challenges in ALR

- **Mechanism**
  - How to handshake?
  - How to re-synchronize at 1 Gb/s
  - How to re-synchronize at 10 Gb/s?

  } How long will it take???
- **Control policy**
  - How to determine when to change rate?
  - How to prevent rate oscillation?
- **Other challenge areas**
  - Effects on packet loss?
  - Interaction with higher layer protocols and applications
    - TCP congestion window?

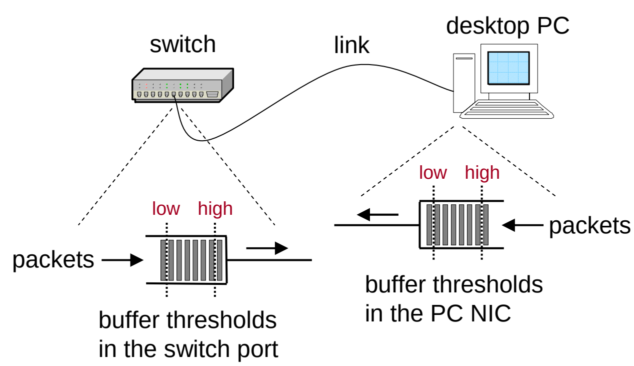
ALR mechanism

- **MAC frame handshake**



ALR control policy

- **When to transition data rate?**
  1) Use queue length thresholds only
  2) Use queue length thresholds and measured utilization

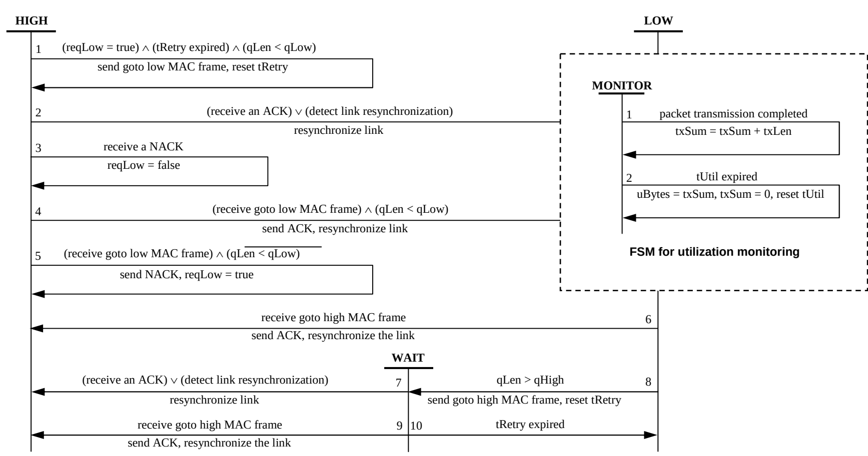


ALR control policy <u>continued</u>

- **Two possible control policies**
  - Threshold
  - Utilization-threshold
- **Threshold policy** ← **Markov model**
  - High and low threshold
  - With an “up timer” and “low timer”
- **Utilization-threshold policy** ← **Simulation model**
  - High threshold to increase rate
  - Explicit low utilization measurement to decrease rate

ALR control policy <u>continued</u>

- **Utilization-threshold policy – detailed FSM**



ALR control policy continued

- **Utilization-threshold policy – simplified**

Executes on receiving a frame …

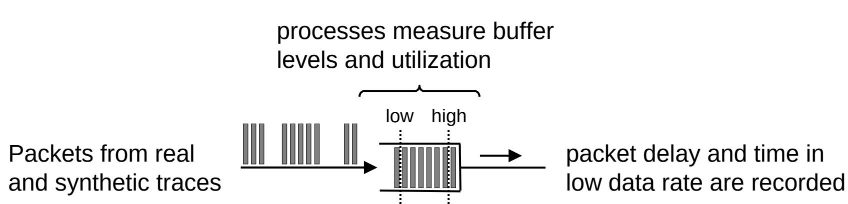
```
if (link rate is low)
  if (buffer size exceeds high threshold)
    handshake for high link rate
```

Executes periodically at end of fixed time period (tUtil) …

```
if (link rate is high)
  if (buffer size less than low threshold)
    if (bytes transmitted less than threshold)
      handshake for low link rate
```


ALR simulation

- **Simulation model used to evaluate ALR**
  - Uses a queue to simulate NIC or switch port output buffer
  - Uses processes for measuring utilization, adapting timers, etc.
  - Implemented in CSIM19
- **Uses a bursty traffic model**
  - Burst sizes are bounded Pareto distributed
  - Inter-burst times are exponentially distributed



ALR simulation continued

- **Model configuration**
  - Single link capable of 1 Gb/s and 100 Mb/s
  - Low threshold is 0KB and high threshold is 32KB
  - Byte transmitted count threshold is 5% of 1 Gb/s
  - Bytes transmitted count is checked once every 10ms
  - Rate switching time is 1ms
- **Traffic model**
  - Based on results and model from Harrison [1]
  - Required utilization determines inter-burst times
  - To generate traces with a mean burst size 8.5KB:
    - Pareto lower bound = 1518 bytes, upper bound = 2.5 GB, index (α) = 1.5
    - Data rate is 1 Gb/s with burst intensity at 80% of data rate
    - Packet size is fixed at 1500 bytes
  - Mean burst size is determined by varying Pareto index
    - For 100KB, α = 0.8; for 2MB, α = 0.5

[1] A. Field, U. Harder, and P. Harrison, "Network Traffic Behaviour in Switched Ethernet Systems", *Performance Evaluation*, Vol. 58, No. 2, pp. 243-260, 2004.

ALR simulation results

- **For ALR utilization-threshold policy**
  - Shows time in 100 Mb/s (low) rate

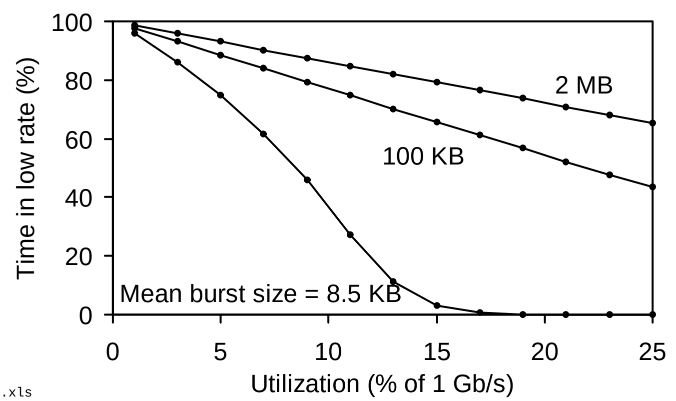


fig2.xls

ALR simulation results continued

- **For ALR utilization-threshold policy**
  - Shows mean packet delay

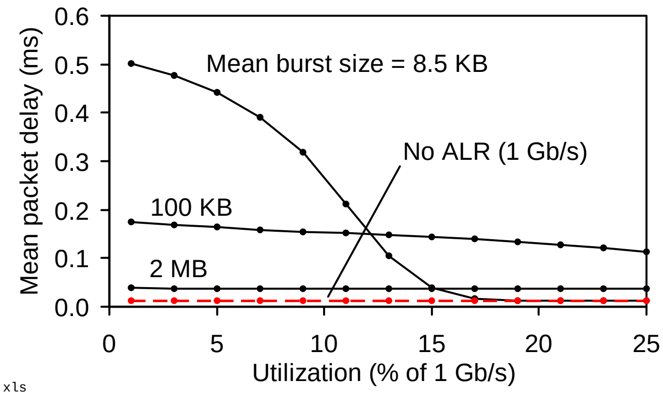


fig3.xls

ALR simulation results continued

- **For ALR utilization-threshold policy**
  - Shows mean packet delay as a function of tUtil
    - This graph only is for Poisson traffic

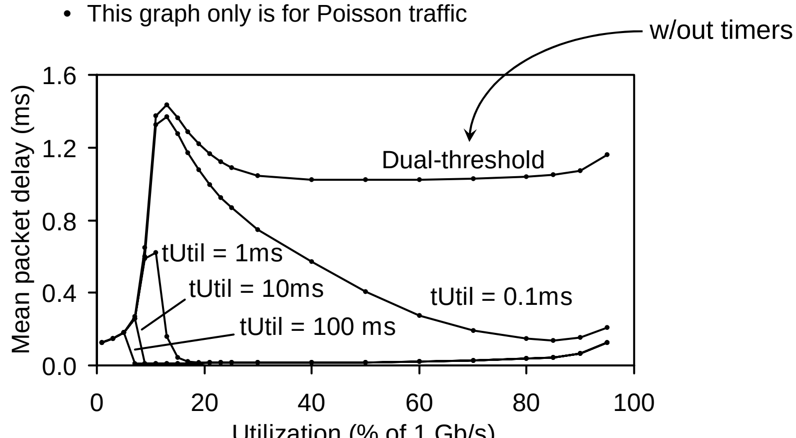


fig10.xls

ALR simulation results continued

- **Poorly tuned utilization-threshold policy**
  - Shows *per packet* delay for a burst from 0.1 to 0.5 s

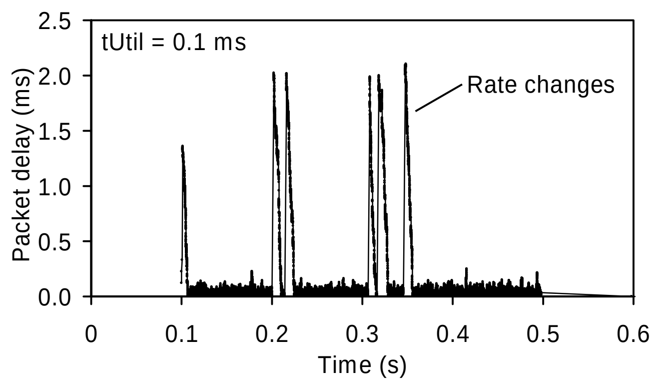


fig11.xls

ALR simulation results continued

- **What is the additional delay for a burst?**
  - ALR can "make it worse" if policy is not carefully designed

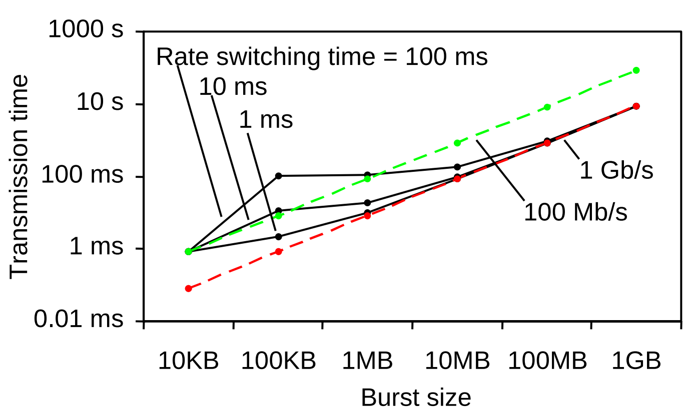
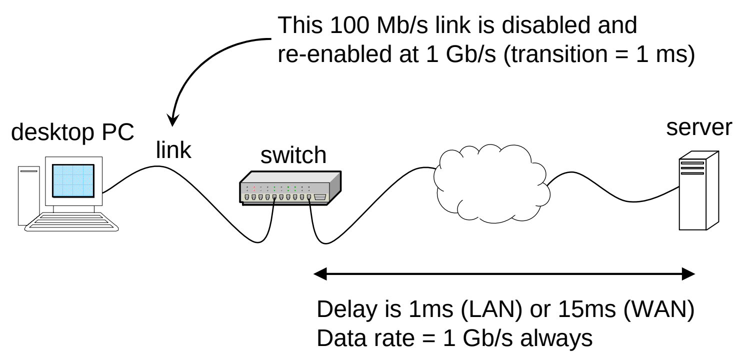


fig6.xls

More ALR simulation

Preliminary results

- **Using ns2 to evaluate effects of ALR on TCP**
  - Study LAN and WAN environments
  - FTP between two nodes via an intermediate node
  - All default except `window_` set to 1000 (default is 20)



More ALR simulation <u>continued</u>

Preliminary results

- **Important note on ns2**

All buffers are destroyed in an ns2 model when a link is disabled/enabled. So, results here are much worse than realistic.

We plan to develop a realistic model.

More ALR simulation continued

Preliminary results

- **ns2 LAN environment**
  - Shows TCP congestion window (transition at Time = 5 s)

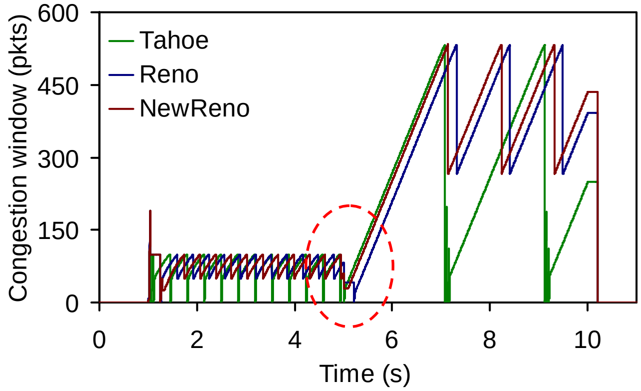


fig7.xls

More ALR simulation continued

Preliminary results

- **ns2 WAN environment**
  - Shows TCP congestion window

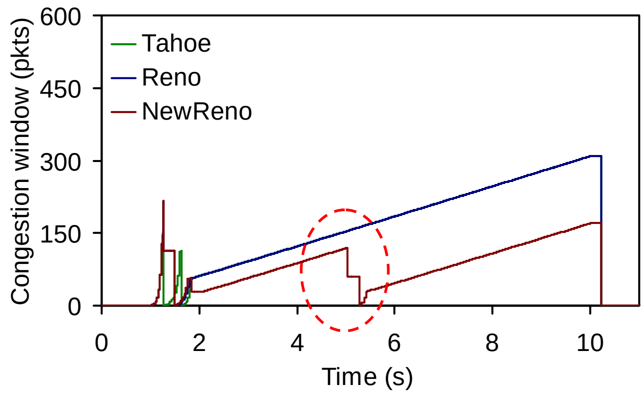


fig8.xls

More ALR simulation <u>continued</u>

Preliminary results

- **ns2 WAN environment**
  - Shows TCP congestion window (with 10 ms transition time)

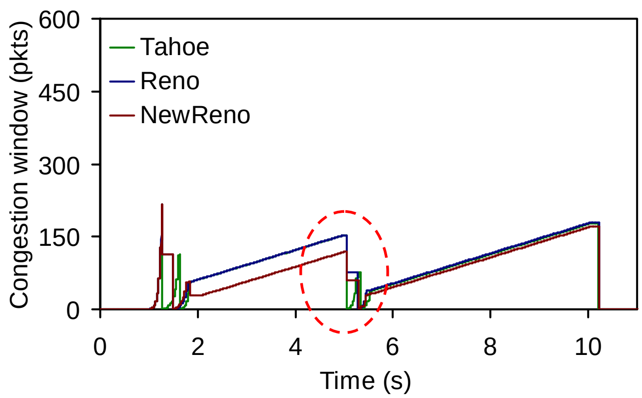


fig9.xls

ALR – a picture tells the story

- **Snapshot of a typical Ethernet link with ALR**

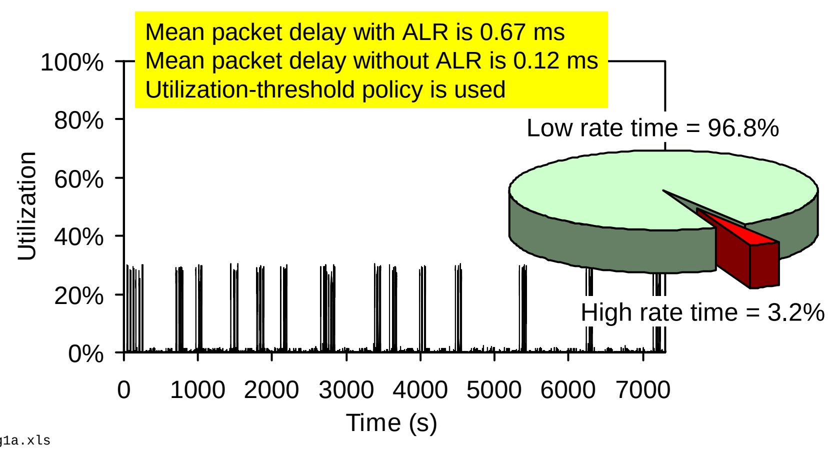


fig1a.xls

Benefits of ALR

- **ALR for 1 Gb/s**
  - Most NICs and most energy to be saved
  - Benefits for homes and offices
  - Battery life benefit for notebooks

  } "Save the planet"
- **ALR for 10 Gb/s (copper)**
  - Reduce power burden in data centers
  - Reduce cooling burden in data centers
  - Possibly increase switch/router port capacity

  } Operational benefits
- **ALR generally…**
  - Demonstrates that network manufacturers and standards organizations take seriously the issue of energy efficiency

Current status of ALR

- **Upcoming CFI at November 802.3 meeting**
  - If CFI passes will have a study group
  - Mechanism to be named to Rapid PHY Selection (RPS)?
    - Addresses mechanism, not policy
- **Chair study group is TBD**
  - Possible name of group: "Energy Efficient Ethernet"
- **Initial list of supporters includes**
  - Next slide…

Current status of ALR continued

David Law	3Com
Brad Booth	AMCC
TBD	Broadcom
TBD	Broadcom
Hugh Barrass	Cisco
Andrew Fanara	Environmental Protection Agency
Joel Goergen	Force10 Networks
TBD	TooBD
Steve Carlson(?)	HSP Design
Petar Pepeljugoski	IBM Research
Ilango Ganga	Intel
TBD	Intel
Mike McConnell	KeyEye Communications
Geoff Thompson	Nortel Networks
Joseph Babanezhad(?)	Plato Networks
George Zimmerman (?)	Solarflare
Shimon Muller	Sun
Rahul Chopra	Teranetics
Bob Noseworthy(?)	University of New Hampshire
TBD	TBD

Initial list of IEEE 802.3 supporters of ALR CFI

+ Claudio De Santi (Cisco)?

Application and savings

Data Center

- Consider a 200 W server with one 10 G copper Ethernet NIC
  - Assume 10 W savings per NIC at each end of link from ALR
  - ALR leverages 10% of total server power
  - Half of time at low link rate saves 5% of total server power
  - ~ $70 million/year for U.S. at 10 million servers

Elsewhere

- Savings on 1 G links
  - Assume 1 W savings each end of link
  - Assume only current stock of Ethernet products
  - ~ $80 million/year

(U.S. only, no power/cooling adders)

Key ALR publications

- H. Anand, C. Reardon, R. Subramaniyan, and A. George, "Ethernet Adaptive Link Rate (ALR): Analysis of a MAC Handshake Protocol", to appear in the *IEEE Conference on Local Computer Networks* in November 2006.
- C. Gunaratne and K. Christensen, "Ethernet Adaptive Link Rate: System Design and Performance Evaluation", to appear in the *IEEE Conference on Local Computer Networks* in November 2006.
- C. Gunaratne, K. Christensen, and S. Suen, "Ethernet Adaptive Link Rate (ALR): Analysis of a Buffer Threshold Policy ," to appear in *IEEE GLOBECOM* in November 2006.
- C. Gunaratne, K. Christensen, and B. Nordman, "Managing Energy Consumption Costs in Desktop PCs and LAN Switches with Proxying, Split TCP Connections, and Scaling of Link Speed," *International Journal of Network Management*, Vol. 15, No. 5, pp. 297-310, September/October 2005.
- M. Bennett, K. Christensen, and Bruce Nordman, "Improving the Energy Efficiency of Ethernet: Adaptive Link Rate Proposal," *White Paper for the Ethernet Alliance*, Version 1.0, July 15, 2006.

Agenda

- **Energy use by IT equipment**
- **Regulatory/policy directions**
- **Reducing *direct* network energy use**
- **Reducing *induced* network energy use**
- **Summary and future directions**

Maintaining connectivity

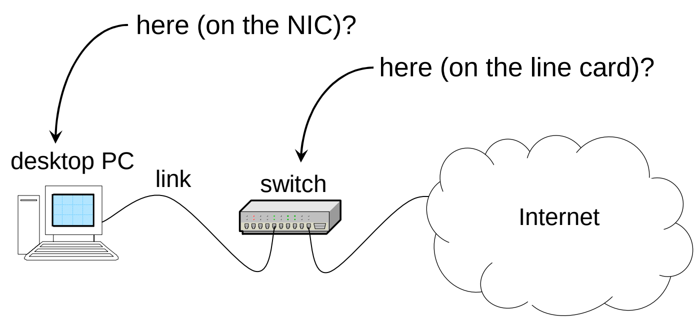
- **Some network applications must keep "connectivity"**
  – TCP connections
  – Other request-response flows
  – Also, ARP, DHCP, and other low level protocols
- **PC must be fully powered-on**
  – Or else "connectivity" is lost and application fails
  – This is **induced** energy use
- **We have studied…**
  – UPnP
  – Gnutella P2P
- **Consumer electronics**
  – Set-top boxes

Maintaining connectivity continued

- **UPnP has a distributed discovery protocol (SSDP)**
  – SSDP requires "always on"
    - Devices send advertise messages
    - Devices wishing to join send discovery messages
- **Gnutella-like P2P servants are always on**
  – Must maintain TCP connections for overlay network
  – Must be able to respond to query messages
    - Queryhit = "I got the file"
  – Must be able to handle HTTP GET

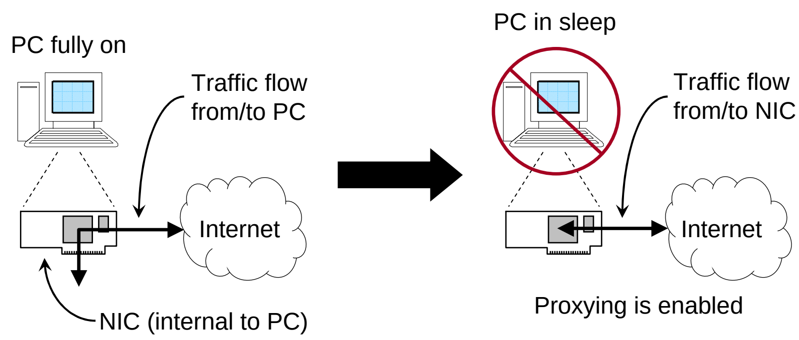
Proxying to keep connectivity

- **What if a lower power subsystem can proxy?**
  - Allow the high-power system (the PC) to goto sleep
  - Wake-up the system when needed (e.g., to download a file)
- **Where to put the proxy?**



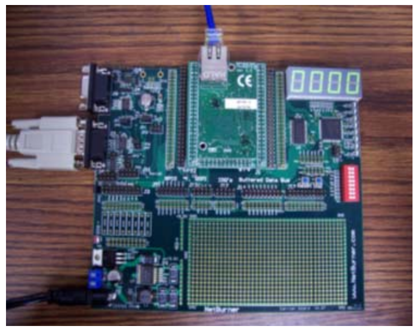
Proxying for P2P

- **We are investigating a "SmartNIC" for P2P**
  - Must add only very little cost to a NIC



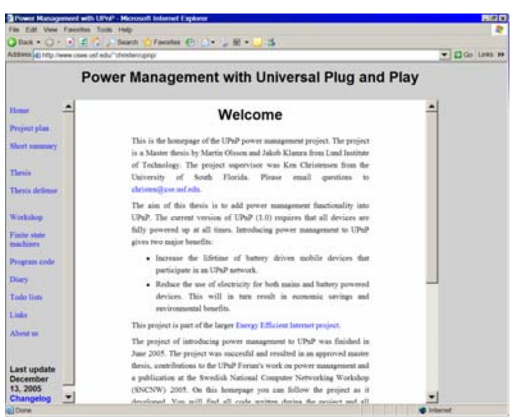
Proxying for P2P <u>continued</u>

- **We almost have a Gnutella P2P proxy running**
  - On a NetBurner embedded Ethernet development kit
  - To be ported to an FPGA-based NIC
    - FPGA work being done by University of Florida



Proxying for UPnP

- **We have developed a proxy server for UPnP**
  - One PC covers for multiple devices
  - UPnP Forum is also studying proxy servers



http://www.csee.usf.edu/~christen/upnp

Application and savings

Data Center
- Not sure there is an application for proxying

Elsewhere
- Much more than ALR
- Hundreds of $millions/year

(U.S. only, no power/cooling adders)

Key proxying publications

- M. Jimeno, K. Christensen, and A. Roginsky, "A Power Management Proxy with a New Best-of-N Bloom Filter Design to Reduce False Positives," to be submitted to the *IEEE International Performance Computing and Communications Conference* in October 2006.
- P. Purushothaman, M. Navada, R. Subramaniyan, C. Reardon, and A. George, "*Power-Proxying on the NIC: A Case Study with the Gnutella File-Sharing Protocol",* to appear in the *IEEE Conference on Local Computer Networks* in November 2006.
- J. Klamra, M. Olsson, K. Christensen, and B. Nordman, "Design and Implementation of a Power Management Proxy for Universal Plug and Play," *Proceedings of the Swedish National Computer Networking Workshop (SNCW 2005)*, November 2005.
- C. Gunaratne, K. Christensen, and B. Nordman, "Managing Energy Consumption Costs in Desktop PCs and LAN Switches with Proxying, Split TCP Connections, and Scaling of Link Speed," *International Journal of Network Management*, Vol. 15, No. 5, pp. 297-310, September/October 2005.

Agenda

- **Energy use by IT equipment**
- **Regulatory/policy directions**
- **Reducing *direct* network energy use**
- **Reducing *induced* network energy use**
- **Summary and future directions**

Summary

- **Network equipment can be "green"**
  - Server community has embraced this direction
- **ALR will reduce direct energy use in links**
  - Open challenges exist at all layers
  - Needs to move to standardized and adopted
- **ALR may enable savings beyond the PHY**
- **Induced energy use at the edge is a big target**
  - Proxying is one direction

Future work

- **Can ALR drive greater power savings in switches?**
  - Enable V/F scaling within line card? Within switch fabric?
  - Are some architectures "better" for this?
- **ALR signaling for multiple parallel links**
- **Sufficient idle periods to fully power down a switch?**
  - What about internal back-up circuitry?
- **What are ALR effects on higher layers?**
  - Effects on throughput and delay

➡ **Would like to address this with a Cisco URP**

Any questions?

Defends in December

\>>> Fresh Ph.D. for hire <<<
Chamara Gunaratne

Did most of the work you just saw

http://www.csee.usf.edu/~pgunarat/

Ken Christensen
Computer Science and Engineering
University of South Florida
Tampa, FL 33620
christen@cse.usf.edu
(813) 974-4761

Bruce Nordman
Energy Analysis
Lawrence Berkeley National Laboratory
Berkeley, CA 94720
BNordman@LBL.gov
(510) 486-7089