ethernet alliance

PAUSE Power Cycle: A New Backwards Compatible Method to Reduce Energy Use of Ethernet Switches

Version 1.0 April 2008

Author:

Francisco Blanquicet
Graduate Student
Department of Computer Science and Engineering
University of South Florida
4202 East Fowler Avenue, ENB 118
Tampa, FL 33620
fblanqui@cse.usf.edu

Graduate Advisor: Dr. Ken Christensen

Ethernet Alliance | 3855 SW 153rd Drive | Beaverton, OR | 97006| USA
www.ethernetalliance.org

Executive Summary

Energy efficient Ethernet switches can provide significant global savings in electricity consumption. The IEEE 802.3az task force (Energy Efficient Ethernet) is exploring methods for scaling Ethernet link rate as a function of utilization to save energy. IEEE 802.3az is studying methods that require both end points of an Ethernet link to implement new mechanisms and policies, thus these methods cannot achieve energy savings with existing Ethernet devices. In this white paper, a new backwards compatible method for achieving energy savings in Ethernet switches is designed, emulated, and evaluated. This new method is called PAUSE Power Cycle (PPC). PPC uses 802.3 PAUSE flow control to cycle Ethernet links between ON and OFF (PAUSEd) states. Major energy savings can be achieved if electrical components are powered-off during the times the link is OFF. Preliminary experimental results show that PPC with a cycle time of 50 ms ON and 50 ms OFF has little or no perceivable effects to users of popular Internet applications such as web browsing, file downloading, real-time video, and playback video (e.g., YouTube). Estimates show that energy savings in the range of $132 million dollars per year in the U.S. can be achieved if PPC is adopted.

1. Introduction

Because of economic and environmental reasons, reducing the energy consumption of IT equipment and other devices connected to the Internet has become a major issue. Most IT equipment is left powered-on 24/7 to provide full and fast connectivity to users. As of 2000, there were approximately 95,000 Ethernet switches deployed in the U.S. alone consuming approximately 3.2 Terawatt hours of electricity [6]. It is well recognized that the vast majority of the links in LAN switches operate at much lower levels of utilization than their data rate [7]. Several consortium-level initiatives are now underway to reduce energy use of IT equipment [13, 15]. Standards initiatives for reducing energy used by networks include IEEE 802.3az Energy Efficient Ethernet [10]. The Ethernet Alliance played a major role in kicking off the EEE effort with the publication of a White Paper on Adaptive Link Rate [9]. In the commercial arena, D-Link is now offering “Green Ethernet” switches that use variable link power as a function of link length and turn off unused ports [14].

The IEEE 802.3az Energy Efficient Ethernet task force is now considering Active/Idle Toggling with Low Power Idle as a means to save energy [11, 12]. The principle behind Active/Idle Toggling is to transmit data as fast as possible and then return to a low power-idle mode. The low-power idle mode consumes minimal power, and during the periods of time in this mode unused circuits are powered off, thus saving energy. IEEE 802.3az also has previously considered Rapid PHY Selection (RPS) as a means of matching data rate with link utilization. RPS was first proposed by Gunaratne and Christensen as Adaptive Link Rate (ALR) [1]. RPS defines a mechanism for switching a link data rate between two Ethernet devices. However, both Active/Idle Toggling and RPS require major changes in the hardware of existing Ethernet devices. Thus, possible energy savings are limited to future Ethernet devices that implement RPS or Active/Idle Toggling mechanisms and policies. In contrast, the PAUSE Power Cycle (PPC) method presented here is a new method that is completely backwards compatible with existing Ethernet devices.

The rest of this white paper is organized as follows. Section 2 describes the PPC method. Section 2.3 describes the estimated energy savings. Sections 3 and 4 describe the evaluation and experimental results. Section 5 describes related work, and section 6 presents conclusions and future work.

2. PAUSE Power Cycle (PPC) – Method Description

PPC is a new method where a LAN switch periodically pauses all active links connected to it and powers off during the times the links are paused. Using IEEE 802.3 PAUSE flow control [8] an Ethernet link can be paused for a specified period of time (i.e., no packets flow on the downstream link of the PAUSE frame sender). It is during the periods in which the links are paused that components in the LAN switch such as PHY, MAC, and interconnects can be powered off, thus allowing for energy savings. The feasibility of implementing PPC, and gaining energy savings from it, are partially motivated from a previous work by Singh and Gupta in LAN switch power management [4], and Rapid PHY Selection (RPS) [1].

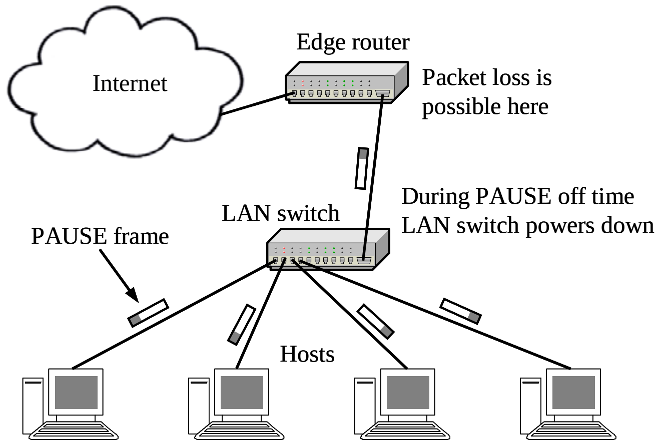

Figure 1— PAUSE Power Cycle (PPG) in a LAN Switch

Figure 1 depicts the proposed PPC method implemented in a LAN switch. Figure 1 shows PAUSE frames being sent on the Ethernet links by the switch to all the devices connected to it. PPC defines two link states:

- ON (link is not paused)
- OFF (link is paused)

In the ON state the link is fully operational, and in the OFF state the downstream interface of the link is paused and no packets can be received on the link.

The PPC method offers a trade-off between performance and energy savings. By periodically pausing the links and powering-off electronic components, major energy savings can be achieved but packet loss may occur due to buffer overflow. In this preliminary work it is shown that although there might be packet loss due to PPC, the overall user experience is not affected and high energy savings are achievable.

2.1 The PPC Finite State Machine (FSM)

PPC can be described by a finite state machine (FSM). In the FSM of Figure 2 the text above the arrow represents the necessary condition to trigger a state transition. The text below the arrow represents the action taken upon a state transition. Two states are defined:

- ON: link fully operational
- OFF: link paused

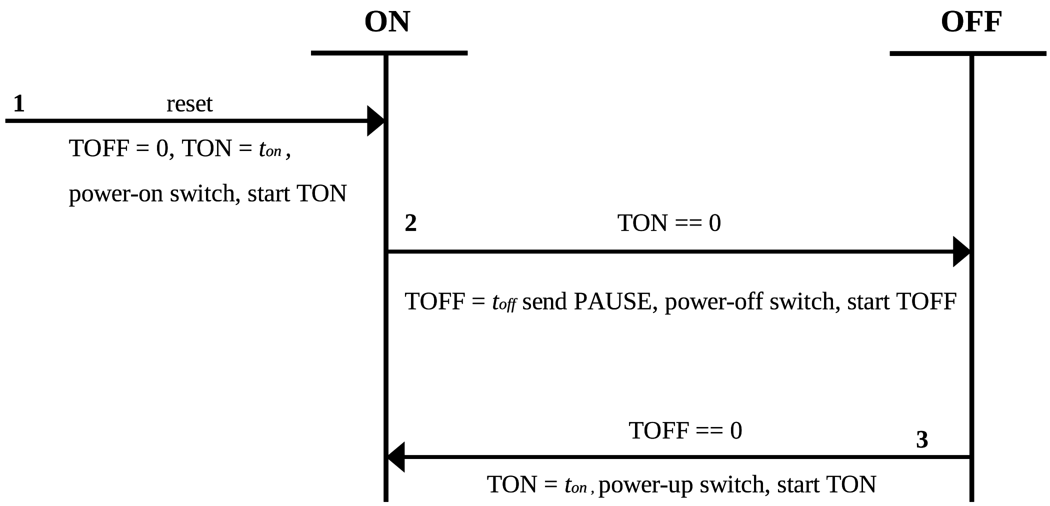

Figure 2— Finite State Machine (FSM) for PAUSE Power-Cycle

Two timers are defined:

- TON: counts time in the ON state
- TOFF: counts time in the OFF state

Two time values are defined:

- t_{on}: amount of time to stay in the ON state
- t_{off}: amount of time to stay in the OFF state

The t_{on} and t_{off} time values can be set by the user as part of an initial switch configuration set-up.

When the switch is powered-on, the timers TON and TOFF are set to the preset time values t_{on} and t_{off}. After presetting the timer values the switch starts the TON timer. When the TON timer expires the switch sends PAUSE frames (with pause time equal to t_{off}) on all its interfaces to stop all incoming traffic. The switch then transitions to the OFF state. While in the OFF state the switch starts the TOFF timer and powers-off components (PHY, MAC, interconnects). When the TOFF timer expires the switch is powered-on, the links are no longer paused, and the switch is back in the ON state. Once back in the ON state the TON timer is again started and the same procedure is repeated. By using IEEE 802.3 flow control to cycle an Ethernet link between ON and OFF states a lower effective data rate is achieved. Since PPC offers the user the ability to set the t_{on} and t_{off} values, any desired trade-off between performance and energy savings can be achieved.

2.2 Setting the t_{on} and t_{off} Values

The values of t_{on} and t_{off} determine the duty cycle of the Ethernet links connected to the switch. Duty cycle is

$$D = \frac{t_{on}}{t_{on} + t_{off}} \quad . \quad (1)$$

Thus, for a 50% duty cycle $t_{on} = t_{off}$. For predefined D and t_{off} the t_{on} value is,

$$t_{on} = \frac{D \cdot t_{off}}{1 - D} \quad . \quad (2)$$

The values of t_{on} and t_{off} are set by the user as an initial configuration step, or could be adaptively set based on utilization. An example of how to adaptively set these values based on utilization is the following: if a switch powers on and no packets are queued for transmission, then the next t_{on} could be shorter in duration and the next t_{off} could be longer in duration. Conversely, if packets are found to be queued, then the next t_{on} could be longer in duration while the t_{off} could be shorter. An adaptive PPC is not discussed here as it is part of a future work.

2.3 Estimated energy savings with PPC

The energy savings achievable with PPC are directly related to the duty cycle and the data rate of the links in a switch. For instance, if PPC is used with a duty cycle of 50%, then the energy reduction of a switch is equal to half the energy needed to power all its link interfaces at the given data rate. Thus, using the estimates from [6] for the year 2000, if all of the Ethernet switches deployed were to work at 50% duty cycle, the annual energy consumption of LAN switches would be reduced to 1.65 Terawatt hours. Taking into account that 1 Kilowatt-hour is approximately 8 cents; savings of about of $132 million dollars per year could be achieved in the U.S. alone.

3. PPC Emulation and Experimental Evaluation

The PPC method was emulated and experimentally evaluated. A test bed consisting of five PCs, an IP-Camera, and a repeater was constructed. Figure 3 shows the test bed configuration. In this test bed the PC labeled “PAUSE Controller” sends PAUSE frames and the repeater propagates them on all other interfaces, thereby emulating a LAN switch which implements PPC. Emulation was used since it is not readily possible to modify a LAN switch to implement the PPC method. With the PAUSE Controller it was possible to vary the t_{off} (time that the link is in the OFF state per cycle) and t_{on} (time that the link is in the ON state per cycle), and thus the link duty cycle. In Figure 3, Q is equal to the t_{off} time, and it corresponds to the quantum (pause time) in the PAUSE frame which is sent to initiate each off period.

The video device used in the test bed was an IP-Camera with a maximum frame rate of 10 images per second. All Ethernet links in the test bed were full duplex 100 Mb/s links.

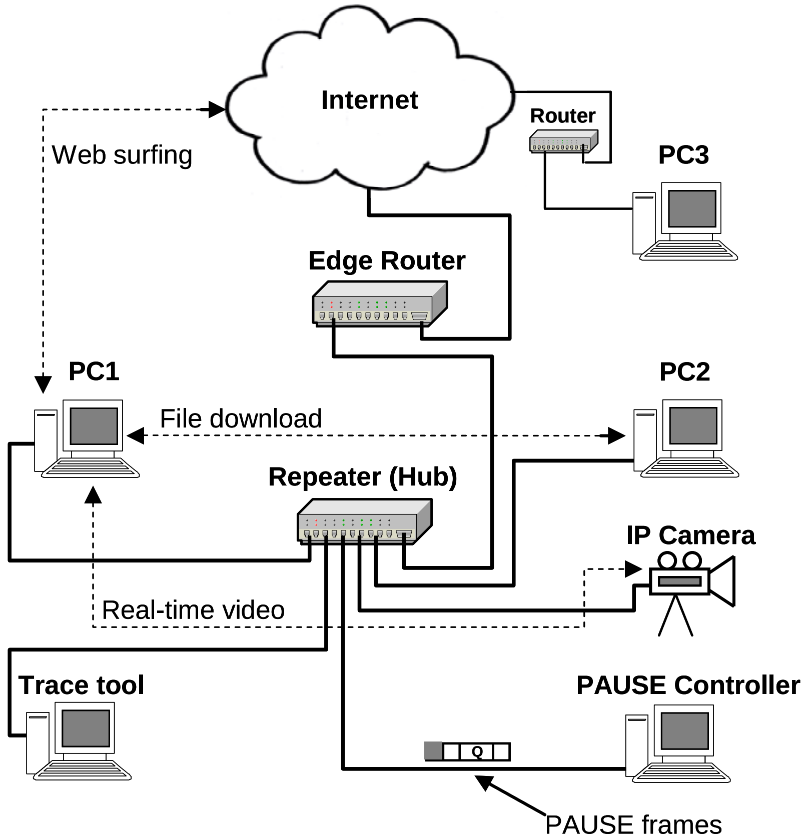

Figure 3 — Test bed of emulated PAUSE power cycle

3.1 Description of Experiments

In order to evaluate the effects of the PPC method on different users of popular Internet applications, a series of experiments were performed. A total of seven test subjects participated in these experiments, and they were all fellow students at the University of South Florida. The experiments were designed to test the user annoyance level caused by PPC controlling the Ethernet link. In the context of this white paper, the user annoyance level is a quantifiable measure that describes the user-perceived quality in the speed of the Ethernet link while it was being controlled by PPC. After each experiment the user annoyance level was measured using a Mean Opinion Score (MOS) scale as follows: (1) bad, (2) poor, (3) fair, (4) good, and (5) excellent. The higher the score, the less annoyed the user was. MOS is the arithmetic mean of the individual scores ranging from 1 (poor) to 5 (excellent), and it provides a numerical indication of the perceived quality of the Ethernet link. The following experiments were defined

- *Web browsing experiment* – test subject freely browses the Internet.
- *File download experiment* – test subject downloads a 512-MB file.
- *Real-time video experiment* – User watches a real-time video of two subjects walking in front of an IP-Camera. This is a simulation of a video-conference setting.
- *Playback video experiment* – User watches a play-back videos (e.g., You-Tube videos)

For each experiment the test subject used PC1 (shown in Figure 3). During each experiment the PAUSE controller (shown in Figure 3) was used to apply the different settings that are described in section 3.2. All settings were applied in a complete random manner.

3.2 Experiment Settings

The control variables for all the experiments were t_{off} and the duty cycle. In each experiment the following t_{off} settings were used: 0 (base case), 300, 100, and 50 milliseconds. Two different settings for the duty cycle were used: 50% and 30% (representing 50 Mb/s and 30 Mb/s PPC data rate). Table 1 summarizes these settings.

Duty Cycle	50 %		30 %	
Data Rate	50 Mb/s		30 Mb/s	
PPC Parameter	t_{on} (ms)	t_{off} (ms)	t_{on} (ms)	t_{off} (ms)
Setting 1	50	50	22	50
Setting 2	100	100	43	100
Setting 3	300	300	129	300

Table 1— Experiment Settings

The response variables were the link throughput or effective data rate of a data transfer application on the PPC controlled link, and the annoyance level as perceived by a human user when using typical applications on the PPC controlled link. The value t_{on} was calculated as described in equation (2). The link throughput was measured using the benchmark Netperf [16]. For each experiment the users started with the base case (no PPC) as this gave them an idea of what the link speed was in the absence of PPC. Then the PPC controller of Figure 3 was used to apply the three t_{off} settings previously described in a complete random manner. The users did not know what setting was being used by PPC during the each experiment run.

4. PPC Experimental Results

Figure 4 shows the averaged results (seven trials for each setting) for the web browsing and file download experiments. Users gave the web browsing experiment with the highest degree of granularity (50 ms of PAUSE time) the highest rating. On the other hand, the web browsing experiment where the PAUSE time was the largest (300 ms) received the worst rating from the users.

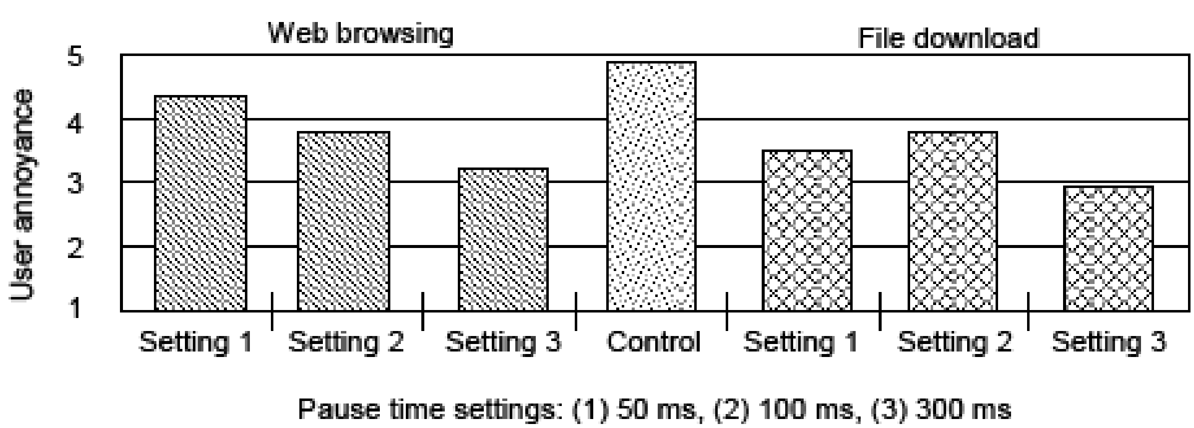

Figure 4 — Web browsing and File download results with PAUSE Power-Cycle

Even though in all three settings the data rate was the same, the 50 ms setting spends less time in the OFF state. With the 300 ms setting users may have experienced lower speeds while web browsing. Web sites with over 50 MB in pictures and graphics were slower to appear, indicating that there was a decrease in the speed of the link. In the case of the file download experiment we see a slight change in the results. Even though with all three settings the files took the same amount of time to download, users perceived the experiment with the 100 ms of PAUSE time as performing slightly better than the 50 ms setting.

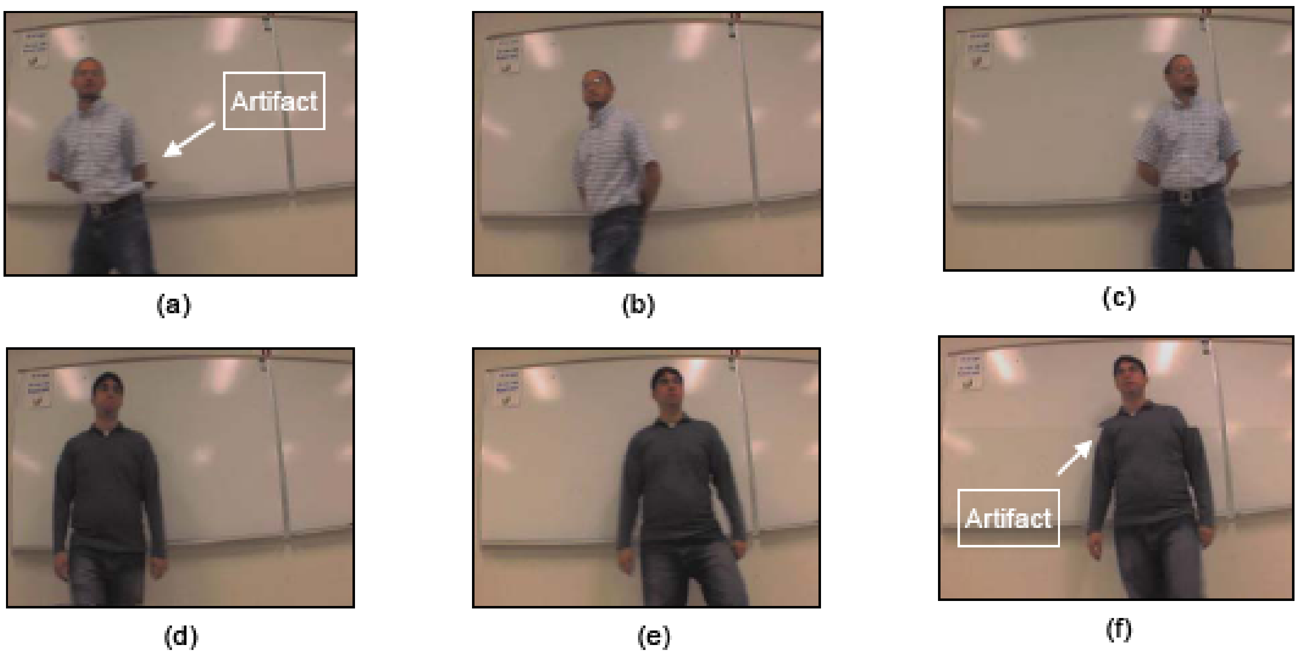

Figure 5— Real time Video Snapshots with ***PPC***

Figure 5 shows six snapshots taken during the real-time video experiment with a duty cycle of 50% and a t_{off} of 300 milliseconds. It can be seen that under this pause time setting the video images present minor artifacts (sliced where the arrows point), Figures 5(a) and 5(f). This happens when the speed and movement direction of the objects in the scene change rapidly. The other PPC settings did not produce any noticeable difference in the quality of the video image when compared to the base case (no PPC). In the play-back video experiment, there was no noticeable difference perceived by the users. This can be explained because these videos are first buffered locally and then played, thus the quality of the video remains the same no matter what PPC setting is being applied to the link.

5. Related Work

In recent years there has been a great deal of interest in reducing the energy consumption of the Internet and Ethernet devices. In [5], the excessive energy use of the Internet was identified, and putting network interfaces, routers, and switches to sleep to save energy was proposed. However, it was also warned that some changes in the current protocols and the architecture of the Internet would be needed to maximize energy savings. In [4], it was shown that during periods of low traffic activity sleeping is feasible for saving energy in LAN switches, but it was suggested that new topologies that allow more ports to sleep would be needed along with hardware that supports sleeping to maximize energy savings and minimize losses. In [3], a new algorithm called Dynamic Ethernet Link Shutdown (DELS) was proposed. The algorithm makes a decision on whether to shutdown a link transceiver circuitry based on buffer occupancy, previous packet arrival times, and a maximum bounded delay. However, it requires the introduction of new MAC frames in order to communicate to the other end of a link transition. Moreover, in order to avoid the auto-negotiation and re-synchronization times in DELS, a modification to the IEEE 802.3 standard is required. The IEEE 802.3az has previously considered Rapid PHY Selection (RPS) as a mechanism for matching data rate to link utilization, but it requires both sides of a link to implement the same RPS mechanisms and policies. Most recently, the IEEE 802.3az is considering Active/Idle Toggling as the method for achieving energy savings, but it may require extensive modifications to the switch hardware and the introduction of a new PHY. In contrast, the PPC method described here does not require the introduction of new MAC frames, thus avoiding hardware reconfiguration; no changes in the topology or the architecture of the Internet are needed, and can be used with existing Ethernet devices already in the market.

6. Conclusions and Future Work

Energy efficient Ethernet devices provide a great opportunity to significantly reduce the electricity consumption of IT equipment and have a global impact on energy savings. This white paper has presented a new method – called PAUSE Power Cycle (PPC) – for reducing the power consumption of Ethernet LAN switches. PPC uses the existing IEEE 802.3 PAUSE flow control to stop incoming transmissions on an Ethernet link. PPC reduces the power consumption of a switch by powering off components in the switch during the times the links are paused. PPC is a new method that is fully compatible with existing Ethernet devices. The PPC method was evaluated experimentally, and a test bed which emulates a real PPC Ethernet LAN switch was built for this purpose. Preliminary experimental results show that the PPC method has little effects on users of popular Internet applications (web browsing, file downloading, and real-time video). Given the large number of Ethernet switches deployed in the U.S., it is estimated that $132 million dollars in electrical energy can be saved in the U.S. alone if 50% of these devices implement the PPC method. An adaptive PPC method that can adapt the duty cycle and t_{off} time based on utilization is future work. Such a method could potentially increase energy savings while decreasing user annoyance.

6. List of References

[1]C. Gunaratne, K. Christensen, and S. Suen, “Ethernet Adaptive Link Rate (ALR): Analysis of a Buffer Threshold Policy,” *Proceedings of IEEE GLOBECOM*, November 2006.

[2] 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.

[3] M. Gupta and S. Singh, “Dynamic Ethernet Link Shutdown for Energy Conservation on Ethernet Links,” *IEEE International Conference on Communications*, pp. 6156-6161, 2007

[4] M. Gupta, S. Grover, and S. Singh, "A Feasibility Study for Power Management in LAN Switches," *Proceedings of the 12th IEEE International Conference on Network Protocols*, pp. 361-371, October 2004.

[5] M. Gupta and S. Singh, "Greening of the Internet," *Proceedings of ACM SIGCOMM*, pp. 19-26, August 2003.

[6] K. Roth, F. Goldstein, and J. Kleinman, "Energy Consumption by Office and Telecommunications Equipment in Commercial Buildings, Volume I: Energy Consumption Baseline," National Technical Information Service (NTIS), U.S. Department of Commerce, Springfield, VA 22161. NTIS Number: PB2002-101438

[7] 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.

[8] Annex 31 B, "Carrier Sense Multiple Access with Collision Detection (CSMA/CD) Access Method and Physical Layer Specification", IEEE standard 802.3, 1998 Edition.

[9] M. Bennett, K. Christensen, B. Nordman, "Improving the Energy Efficiency of Ethernet: Adaptive Link Rate Proposal," *Ethernet Alliance White Paper*, URL: http://www.ethernetalliance.org/technology/white_papers/alr_v10.pdf

[10] IEEE 802.3 Energy Efficient Ethernet Task Force, 2007. URL: http://www.ieee802.org/3/az/index.html

[11] R. Hays, "Active/Idle Toggling with OBASE-x for Energy Efficient Ethernet," IEEE 802.3 Energy Efficient Ethernet Task Force, November 2007 Presentation, URL: http://www.ieee802.org/3/az/public/nov07/hays_1_1107.pdf

[12] R. Hays, "Active/Idle Toggling with Low-Power Idle," IEEE 802.3 Energy Efficient Ethernet Task Force, January 2008 Presentation, URL: http://www.ieee802.org/3/az/public/jan08/hays_01_0108.pdf

[13] Climate Savers Computing, 2007. URL: http://www.climatesaverscomputing.org/

[14] D-Link Innovative Green Ethernet, 2007. URL: http://www.dlink.com/corporate/environment/dlink-green-ethernet/

[15] The Green Grid: Improving Power Efficiency in Data Centers, 2007. URL: http://www.thegreengrid.org/home

[16] Netperf, "Network Performance Benchmark," URL: http://netperf.org

[17] YouTube, 2008. URL: http://youtube.com.