G Model
SUSCOM-199; No. of Pages 14
ARTICLE IN PRESS
Sustainable Computing: Informatics and Systems xxx (2017) xxx–xxx
Contents lists available at ScienceDirect
Sustainable Computing: Informatics and Systems
journal homepage: www.elsevier.com/locate/suscom

Dual-Mode Energy Efficient Ethernet with Packet Coalescing: Analysis and Simulation

Mehrgan Mostowfi[a,*], Khalil Shafie[b]

[a] Department of Accounting and Computer Information Systems, Monfort College of Business, University of Northern Colorado, CB 128, 501 20th St, Greeley, Colorado 80369, USA
[b] Department of Applied Statistics and Research Methods, College of Education and Behavioral Sciences, University of Northern Colorado, CB 124, 501 20th St, Greeley, Colorado 80369, USA

ARTICLE INFO

Article history:
Received 11 November 2016
Received in revised form 18 July 2017
Accepted 20 October 2017
Available online xxx

MSC:
00-01
99-00

Keywords:
High capacity Ethernet
Frame coalescing
M/G/1 queue
Energy efficiency
IEEE 802.3bj
Performance evaluation

ABSTRACT

The recently-proposed and studied Dual-Mode Energy Efficient Ethernet (EEE) amends the conventional EEE with an additional low-power mode called Fast Wake. Dual-Mode EEE has shown to be much more effective than EEE in reducing the power consumption of Ethernet links at 40 Gb/s and higher due to impractically-high transition times of using only one low-power mode at these capacities. However, even with two low-power modes, the still relatively-high transition times between the modes prevent Dual-Mode EEE from taking full advantage of the power reduction of the low-power modes and approach energy-proportional operation.

We design and formulate Packet Coalescing for Dual-Mode EEE links at 40 Gb/s and higher. Packet Coalescing is to collect multiple packets before transmitting them on a link as a burst of back-to-back packets. By coalescing packets into bursts, the overhead of transition times can be reduced and nearly energy-proportional operation can be achieved, albeit at the expense of a bounded increase in the packet delay. We analytically model the power consumption of Dual-Mode EEE with Packet Coalescing and present closed-form formulae for the average percentage of time a link would spend in each low-power mode and transition. We also simulate a 40 Gb/s Dual-Mode EEE link with Packet Coalescing and present results suggesting that the compromise in the average per-packet delay is likely to be worthwhile under most circumstances considering the power reduction.

1. Introduction

Conventional Ethernet links stay fully powered on all the time regardless of whether there is a packet in the link's buffer to transmit/receive or not. However, it is a known fact that there is a high number of idle periods for an Ethernet interface where there is no activity on the link. Even in environments with seemingly very high traffic load such as datacenter links, research has shown that the typical utilization of Ethernet links is surprisingly between 10 and 50% [3,4,17]. Energy Efficient Ethernet (EEE) is a standard to reduce the energy consumption of Ethernet links by enabling the link to power down [10].

EEE defines a low-power mode called the Low-Power Idle Mode (LPI) for an Ethernet link. In the LPI mode, the physical layer of an Ethernet interface is powered off and some elements in the receiver are stopped. As a result, the power consumption of the link is reduced to a fraction of its peak consumption.[1] While no explicit policy of when to enter or exit LPI is devised in the standard, a widely-used policy of utilizing LPI is to enter LPI when there are no packets to transmit and return to normal operation – the Active mode – once a packet arrives for transmission from upper layers. It is estimated that implementing EEE can result in significant energy savings – in the order of about $180 million per year if it were implemented in all current 1 Gb/s edge links in both residential and commercial buildings [6].

With the above transitioning policy, it has been shown that EEE could suffer from an inefficiency caused by high transition times especially in cases where packets arrive to the link in single-packet

* Corresponding author.
E-mail addresses: mehrgan.mostowfi@unco.edu (M. Mostowfi), khalil.shafie@unco.edu (K. Shafie).

[1] Note that it may not be correct to use the verb "consume" for power. However, power and energy are sometimes used interchangeably in the context of Green Computing, which is, strictly speaking, incorrect. Energy is the work done in a system to generate the desired outcome (transmission of bits, calculation, etc.), while power is the rate at which energy is consumed (or produced). While special care has been taken in this paper not to confuse the two concepts, the verb "consume" is sometimes used for both power and energy, as is common in the literature. The term "power draw" is the correct term.

https://doi.org/10.1016/j.suscom.2017.10.008
2210-5379/

or small batches [23]. In these cases, the link wakes up to transmit a single packet and immediately transitions back to LPI, with each transition taking multiple times that of a packet transmission while consuming almost the same power as the link's peak power, resulting in the link consuming not less than 80% of its peak power consumption even under very light traffic loads. An alternative transitioning policy is to coalesce the outgoing packets into bursts, consequently decreasing the number of necessary transitions to one per burst. This method, called Packet Coalescing, was introduced as an addition to EEE in 2010 [6,25] and was shown to be able to achieve nearly energy-proportional operation for EEE links. Packet Coalescing was inspired by a similar method called Receive-Side Coalescing which had been in use in many high-speed Ethernet interfaces – mostly on the receive side – to reduce CPU overhead for packet processing [15].

EEE was originally developed for 100 Mb/s, 1 Gb/s, and 10 Gb/s Ethernet, but not for any higher capacities. After the standardization of 40 Gb/s and 100 Gb/s Ethernet in 2010, an optional EEE capability was added to it in 2014. EEE for 40 Gb/s and 100 Gb/s differs from EEE for 10 Gb/s and lower in that it contains an additional mode called Fast Wake. The reason to include this additional mode was to address the even longer transition times in the 40 Gb/s and 100 Gb/s EEE. In its current form, the new Fast Wake mode is not a low-power mode. However, as it will be argued here and has been in other publications and presentations cited later, it is very likely that the new Fast Wake mode could act as an intermediate low-power mode and enable additional energy savings for 40 Gb/s EEE and higher. Besides feasibility, the need for Fast Wake becoming a low-power mode is clearly demonstrated in [2] where simulation results suggest that EEE with a single low-power mode and simple transitioning policy as described above would not yield significant savings, and would also impose a higher average per-packet delay than Dual-Mode EEE. Nonetheless, even with two low-power modes, a simple policy like what is described above would yield limited power savings which would still be far from energy-proportional.

In this paper, we explore Packet Coalescing for Dual-Mode EEE in 40 Gb/s Ethernet and higher [18]. The contributions of this paper are the following:

1. Development of a Packet Coalescing method for Dual-Mode EEE.
2. Development of an analytical model for the average power consumption of Dual-Mode EEE with the new Packet Coalescing policy using a Markov chain.
3. Development of a simulation model of Dual-Mode EEE with Packet Coalescing using the CSIM simulation library in C.
4. Performance evaluation of this method using the developed models.

2. Related work

The first simulation study of EEE was performed in the pioneering work of Reviriego et al. [23] in 2009 where the inefficiency of EEE caused by the relatively-long transition times was explored. The conclusion was that the transition times are too high in comparison with the time needed to transmit one frame, reducing the energy efficiency of EEE significantly. Later, an analytical model for the average power consumption of a EEE link was developed in 2011 [16]. As was proposed as future work in [23] to "group frames before waking up the link", the authors proposed and studied Packet Coalescing for EEE by simulation in [6,25]. Analytical models for Packet Coalescing for EEE were published in the following years in [9,20,1].

Even though the EEE standard for 40 Gb/s and 100 Gb/s Ethernet as defined in IEEE 802.3bj [12] contains two modes, Deep Sleep and Fast Wake, the Fast Wake mode is **not** a low-power mode. In Fast Wake, the link would consume 100% of the link's full power consumption just like normal operation and is just defined to keep the sender and receiver in alignment while the receiver transitions to Deep Sleep. Defining two modes of low-power operation – effectively turning Fast Wake into a low-power mode – was first introduced as a potential option for 100 Gb/s EEE in a presentation at an IEEE 802.3bj meeting in 2012 [2]. In this presentation, a limited performance study of EEE with two modes with burst transmission of packets where a few packets are coalesced and sent as a batch on the link was presented. The main focus of this presentation was to show the possibility of this approach, not a detailed performance evaluation. Consequently, the simulation model was simplistic, the coalescing method was not documented and the results were preliminary.

The first thorough study of Dual-Mode EEE was performed in 2015 which used simulation to investigate the power-delay tradeoffs of the method [19]. The study in [19] showed that one low-power mode for EEE at 40 Gb/s and higher would be inefficient and more than one mode is needed. It also showed that even with two low-power modes, the power consumption would still be far from energy-proportional. Packet Coalescing for Dual-Mode EEE was first proposed shortly after the publication of [19] in 2015 [18]. The average power consumption and per-packet delay caused by Dual-mode EEE was analytically modeled in 2016 [21,22]. In [21], an analytical model for the average fraction of time a Dual-Mode EEE link would spend in each mode (low-power modes and transitions) under Poisson traffic and various traffic load was developed. In [22], a model for the average packet delay for the method under the same traffic and parameters assumption was developed. The current paper builds on [18] mainly by adding an analytical model for the average power consumption of Dual-Mode EEE with Packet Coalescing and extending the experiments and results to further evaluate the performance of the method.

3. Dual-Mode EEE

In 2008, the IEEE 802.3 working group determined that capacities higher than 10 Gb/s were needed to address the growing demands of packet-switched networks, and formed the IEEE P802.3ba 40 Gb/s and 100 Gb/s Ethernet Task Force [11] to draft a standard for Ethernet at these capacities. The final result of this effort was the amendment of the IEEE Std 802.3ba 2010 40 Gb/s and 100 Gb/s Ethernet to the IEEE 802.3 Ethernet standard in 2010. In another amendment to the standard, the IEEE Std 802.3bj [12], the optional EEE capability for 40 and 100 Gb/s Ethernet was defined.

EEE for 40 Gb/s and 100 Gb/s Ethernet as defined in IEEE 802.3bj contains a Deep Sleep mode which is identical to LPI. In addition to Deep Sleep, a second mode called Fast Wake is also defined which does not yield any power savings; this mode is only used to keep the sender and receiver in alignment while the receiver transitions to Deep Sleep, and is practically the same as the Active mode. The reason for introducing Fast Wake was that with only one LPI mode the transition times to and from the mode would have been impractically high due to the long times needed for removing and reapplying power, as well as realigning the line.

Even though Fast Wake is not defined as a low-power mode, only minimal functionality of the interface is needed in this mode to continue sending LPI signals. So, some components can potentially be turned off or powered down. As suggested in [2], the second low-power mode can be without clock stopping yielding limited power reduction, while having a short transition time compared to Deep Sleep. The changes in the circuitry of an EEE interface to accommodate two low-power modes is beyond the scope of this paper and needs a separate research effort. However, based on the discussions in [2] and the definition of Fast Wake in [12], reducing the