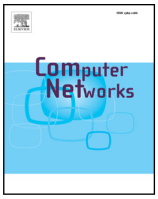
Computer Networks 122 (2017) 217–230

Contents lists available at ScienceDirect

Computer Networks

journal homepage: www.elsevier.com/locate/comnet



Modeling and understanding burst transmission for energy efficient ethernet

Jinli Meng, Fengyuan Ren*, Chuang Lin

Tsinghua National Laboratory for Information Science and Technology, Department of Computer Science and Technology, Tsinghua University, Beijing, 100084, China

ARTICLE INFO

Article history:
Received 8 July 2015
Revised 17 August 2016
Accepted 25 March 2017
Available online 27 March 2017

Keywords:
Burst transmission
Energy Efficient Ethernet
Markov model
Energy saving
Packet delay

ABSTRACT

Recently, the energy consumption of Ethernet has become one of the hottest topics focused by both academic committee and industry, especially with the increase of the link speed from 1Gbps to 10/40Gbps nowadays or even 100/400Gbps in the near future. To save the energy consumed by the Ethernet, the Energy Efficient Ethernet (EEE) is developed and standardized by the IEEE 802.3az work group. When there is no incoming traffic, the EEE can saves 90% of its energy consumption by entering into the Low Power Idle (LPI) mode. To maximize the energy saving of Ethernet, the Burst TRansmission (BTR) mechanism, which defines a new way to utilize the LPI mode, is developed as a mechanism for EEE. Prior work theoretically shows that the BTR mechanism makes a tradeoff between the energy saving and the queuing delay. However, the traffic pattern, on which the performance of EEE greatly depends, is assumed to be deterministic in their analyses. Besides, their models made estimation for many situations. In this paper, we model EEE with the BTR mechanism and provide analytical understanding on the BTR mechanism. We propose two actual models: one focuses on the buffer size limit, the other concentrates on tolerable packet delay additionally. We draw some guidelines of parameter selection and policies design for EEE from combination of theory conclusions and simulation results. The results show that the saved energy can be constrained by link occupancy even though the buffer size is variational. The buffer full triggered wake-up policy can achieve ideal ratio of energy consumption and arrival rate within the scope of the buffer as well. However, the tolerable delay cannot be guaranteed by any policies. The buffer size is even fixed, which affects the flexibility of demanded delay for different business. The policy considering tolerable delay is supposed to be a little better than the other policy, with a little more complicated design. Thus we design an adaptive mechanism both in quantitative and qualitative: detect the load utilization, apply the buffer full triggered wake-up policy for higher load utilization link, while applying the buffer full and timeout triggered wake-up policy for the delay sensitive business and tiny arrival rate.

1. Introduction

Ethernet is widely deployed from 1980s. It is not only the main structure of Data Center Network (DCN), but also the most used access network in the world. The power consumption of Data Center is 2.3% of total power consumption in US [3]. With the growth of the Data Center scale, its power consumption doubles every five years [21]. Among these power consumption of IT infrastructure in Data Center, the network equipment consumes 20% approximately, which cannot be ignored [1,5]. Moreover, the increase of the link speed leads to great power demand. For instance, the Network Interface Cards (NIC or interface) consumes 0.5W for 1Gbps link, while the number is 5W for 10Gbps [9]. The 10Gbps Ethernet is the main trend, 40Gbps Ethernet appears recently. The speed will reach 100/400Gbps in the near future [12,23]. No matter from the angle of environment sustainable development, or from the view of reducing the costs of power consumption, lowering the energy expenditure of the Ethernet devices is imperative.

The average load factors of the Ethernet is low in most time. The figure is 5% for the general computers, up to 30% for busy servers [2,18,19]. Therefore, reducing the power consumption of idle interface can save energy effectively. A norm called Energy Efficient Ethernet (EEE in short) absorbs such a mechanism of reducing power consumption in Ethernet network equipments and hosts during periods of low link utilization. It is issued by the IEEE as a related industry standard, and officially approved just in September 2010. Nowadays, the manufacturers, such as HP, Broadcom and

---

* Corresponding author.
E-mail addresses: mengjl08@csnet1.cs.tsinghua.edu.cn (J. Meng), renfy@tsinghua.edu.cn, renfy@csnet1.cs.tsinghua.edu.cn (F. Ren), chlin@csnet1.cs.tsinghua.edu.cn (C. Lin).

http://dx.doi.org/10.1016/j.comnet.2017.03.020
1389-1286/© 2017 Published by Elsevier B.V.

Asus, have their productions supporting the EEE standard. Some switch manufacturers integrate a variety of power-saving features across some models, and EEE is applied as one of them, such as the Cisco 500 series stackable managed switches [13], the Cisco Catalyst 4500E Switches [15], the Huawei S1700 series switches [14]. Intel is also integrated EEE in their products, such as Intel Ethernet Controller I350 and Intel 82579LM Gigabit Ethernet Connection [15].

EEE provides the mechanism of saving energy by powering off the unused elements of the interface when no transmission is required [6,15]. Then the interface is in a low level of power consumption, which is called Low Power Idle (LPI) mode. Comparing with the normal transmission mode, the interface consumes only 10% energy of that in LPI mode. The normal work state is called active. The states of the interface can be active or LPI. From active to LPI, it takes some time to power off some elements while powering on them during the opposite transition. The sender decides to powered off or powered on and signals the other end of the link during the two transition periods. These two periods are called sleep and wake-up respectively. The standard also provides the protocol of coordinating the transitions between active and LPI.

However, EEE only supplies the mechanism of saving energy, but the details of state transitions are not supplied in the final standard. Two mechanisms of them are more practical and widely used [4,7,8,18]. The first one is **frame transmission** (**FTR**). Its main idea is to wake up the interface immediately once a new packet arrives in the LPI mode. In contrast, when a new packet arrives while the interface is in the LPI mode, it can be stored in the buffer until the buffer goes full. Then the interface is going to recover to the active state. The later mechanism is called **burst transmission** (**BTR**) or **packet coalescing**.

For the implementation of the BTR mechanism, various of policies are designed. When a new packet arrives while the interface is in the LPI mode, it can be stored in the buffer until the buffer contains enough packets. Then the interface is going to be waked up. The wake-up time depends on the buffer occupancy as well as the waiting time of the packets in the buffer. Thus the events of triggering wake-up operation include buffer occupancy and timeout of buffered packets. The first wake-up policy, which is called *buffer full triggered wake-up*, depends on the preset buffer size for the BTR mechanism. The second wake-up policy, which is called *buffer full and timeout triggered wake-up*, depends on both the preset buffer size and the minimum value of maximum delay of buffered packets. Different categories of the flows make different tolerable delays, which are also the maximum values for the specific category. The minimum one among all the maximum delays of the buffered packets can be set as the delay characteristics of the interface. This is the minimum value of maximum delay.

In this paper, we propose analytical models of the BTR mechanism and make comparison with FTR. FTR mechanism could be considered as a special case of BTR when buffer capacity for the LPI state is 0. Thus we only propose the general models and explain the performance analysis individually. The tradeoff between saved energy and performance degradation is discussed both for FTR and BTR. How to choose the practical parameters for advanced deployment of EEE is also advised in this paper. Based on the analysis of the models, we design policies of BTR mechanism for actual network environment.

The rest of the paper is organized as follows. In Section 2, we give more details and descriptions of the FTR and BTR mechanisms, as well as introduce more motivations of our work. Besides, we refer to related works, compare them with our work in this section as well. Sections 3 and 4 describe and solve the analytical models of these mechanisms, one model considers the buffer overflow, while the other considers packets' delay extra. Some practical conclusions are discussed according to the models. Further discussion are represented, we also draw some qualitative guidelines of policy design for EEE from comprehensive of theory conclusions and analysis results. We validate the models by comparing the analytical results with that obtained by means of simulation, make lots of detailed quantitative analysis of energy saving in Section 5. We discussion the extension of EEE models in different traffic pattern and mixture of delay tolerant and delay sensitive traffic in Section 6. Concluding remarks are given in Section 7.

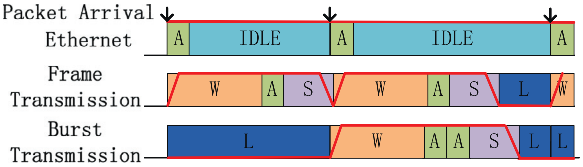


Fig. 1. An event line for Ethernet, frame transmission and burst transmission of EEE.

2. Background and related work

EEE for different speed links has different realization techniques [10,18]. Firstly, 100Mbps and 10Gbps links can be powered off in unidirectional links, while 1Gbps links should be powered off when both directions have no traffic. Secondly, transitions from the Active state to the LPI state cannot be interrupted in 10Gbps links, while immediate activation is caused by arriving packets when the interface is in sleep period for 100Mbps and 1Gbps links. Moreover, EEE for 10Gbps links becomes more and more widely used, the technology is more typical meanwhile. Therefore, we focus on EEE for 10Gbps links only.

As mentioned in Section 1, the interface of EEE has four operation states.

- The state that the interface is transmitting packets is called **Active**, or **A** in short. In the Active state, when the buffer becomes empty and no other packet comes, the interface will take a **Sleep** (or **S** in short) operation.
- The Sleep process lasts for a period T_s. When a packet comes, the interface will continue to executive the Sleep operation; at the end of the Sleep operation, the interface will be waked up immediately. Or else, the interface will enter into the LPI state. The power consumption in the Sleep process is almost the same as that in the Active state.
- The state when the interface has been turned off is called **LPI**, or **L** in short. In the LPI state, the buffer is empty and no packet can be transmitting. Once a packet comes, the interface will start to be waked up. Besides, because the interface is turned off, the power consumption is 10% of that in the Active state.
- The **Wake-up** (or **W** in short) operation lasts for a period T_W; after the interface gets ready to transmit packets, the interface will turn into the Active state. In the Wake-up process, the power consumption is almost the same as that in the Active state.

EEE can save energy by keeping in the LPI state. Assume such a periodic traffic pattern: the first packet comes, the second packet arrives after 8.478 µs, then the third one arrives 11 µs later, which is shown in Fig. 1. Obviously, the time length of each period is 19.478 µs. Without loss of generality, assume that all the packets are of length 1500 bytes. Accordingly, the time of transmitting a packet is 1.118 µs for 10Gbps link.

In traditional Ethernet, the packet can be transmitted immediately when it arrives. Moreover, traditional Ethernet consumes the same power consistently no matter the link is Active or IDLE. The interval of the packets arrival is larger than the transmitting time,