Enhanced Synchronizing Packet Coalescing Mechanism for Improving Energy Efficiency in Ethernet Switch

Young-Hyun Kim[1], Sung-Keun Lee[2,†] and Jin-Gwang Koh[1]

[1]*Department of Computer Engineering, Sunchon National University, Korea*
[2]*Department of Multimedia Engineering, Sunchon National University, Korea*
yhkim31@empas.com, sklee@sunchon.ac.kr, kjg@sunchon.ac.kr

Abstract

As network equipment consumes much more energy than ever before, research effort has been focusing on developing the access network protocol that aims to achieve communication with maximum energy efficiency. Ethernet is the most widely deployed access network protocol around the world. To improve energy efficiency in Ethernet protocol, IEEE 802.3az working group has defined low-power idle(LPI)-based Energy Efficient Ethernet (EEE) specification. In this paper, an enhanced synchronizing packet coalescing mechanism is proposed which improves energy efficiency in small Ethernet switch with basis on EEE specification. This mechanism is designed to predict traffic characteristic for active period by measuring incoming traffic amount during certain period. Once the traffic characteristic is predicted, threshold value would be adjusted to be optimal for the traffic load. Network performance was evaluated through simulation experiment that was performed assuming Poisson distribution traffic and burst traffic pattern respectively. Performance evaluation results indicate that the proposed mechanism improves overall performance compared to traditional mechanism, since it significantly reduces energy consumption rate, even though average packet delay increases a little bit.

Keywords: *Synchronizing packet coalescing mechanism, Energy Efficient Ethernet, IEEE 802.3az, Small Ethernet switch*

1. Introduction

Recently, demands for real-time multimedia service based on Internet have been growing high. To meet those demands, various access network equipments such as Ethernet switch, network interface card as well as core router have developed offering fast data transmission rate. At the same time, an amount of power consumed by network equipment is also growing fast. For this reason, significant amount of research effort has been focusing on the area that aims to save energy consumed by access network equipment [1]. It is reported that Internet core network consumes power more than 6TWh per year. In US, an amount of power consumption by network equipment accounts for about 2% of total power consumption. This amount corresponds to more than 74 TWh per year, which is also said to be 6 trillion dollars [2]. Therefore it is natural for governments and industries to have interest in an efficient use of energy and a design for achieving minimized energy consumption in network system [1]. Ethernet is the most widely deployed access networking technology around the world. With the use of energy efficient protocol on Ethernet, about 3 TWh per year can be saved from wasting. Furthermore, it is expected that energy consumption costs would be reduced by $400 million per year only in U.S. and over $1 billion per year around the world [3]. To this end,

† Corresponding Author

IEEE 802.3az working group set out the formation of the standard for Energy Efficient Ethernet (EEE) since early 2007. On September in 2010, standard applying low power idle (LPI) mechanism was released. Detailed specification adopted to physical layer also was published [4]. LPI is a mechanism that allows network equipment to save power by switching state of Ethernet link into low-power mode when there is no data traffic to send [5, 6]. EEE can be applied to Ethernet switch as well as Ethernet network interface card. By doing this, unnecessary power consumption can be saved in the network environment where network load is low [7]. In this paper, enhanced synchronizing packet coalescing mechanism is proposed which aims to improve energy efficiency in small Ethernet switch by examining characteristic of incoming data traffic. Based on the analysis for the incoming traffic, the proposed mechanism adjusts threshold value in an adaptive way. This mechanism is designed to predict traffic characteristic for the coming period by measuring incoming traffic amount during certain period. Once the traffic characteristic is predicted, optimal threshold value can be determined to be suitable for the traffic load.

The remainder of the paper is organized as follows. Section 2 reviews previous works related to energy control mechanism used in Ethernet switch. Section 3 investigates the problem residing in adaptive PPSE mechanism. To overcome the problem, an enhanced synchronizing packet coalescing mechanism is proposed. Section 4 evaluates performance through simulation. Finally, the paper is concluded in Section 5.

2. Related Work

Periodically paused switched Ethernet (PPSE) was previously proposed to enhance energy efficiency in small-sized Ethernet switch [7]. In this mechanism, port of switch enters either ON state or OFF state. When the port goes to ON state, the switch can send packet normally. When OFF state, power of link turns off, so that packet transmission is stopped. Switch is changing state ON and OFF alternatively while operating. Duty cycle is a value that indicates percentage of time the switch spends in ON state. It can be obtained from (equation 1). TON denotes the time that switch spends in ON states during one interval, which is tightly related to energy consumption rate of switch. TOFF is the time that switch spends in OFF states.

$$D = TON / (TON + TOFF) \quad (\mathbf{1})$$

This mechanism is a solution to achieve significant energy saving at the expense of a acceptable level of packet delay, only if duty cycle, TON, and TOFF are set to be proper value, and incoming traffic load is maintained not exceeding certain level. On the other hand, if traffic load increases over certain level, average packet delay also grows substantially high. For this reason, this mechanism is not suitable for transmitting real-time multimedia data properly. To overcome this challenge, adaptive PPSE mechanism was suggested. While the traditional mechanism has a fixed time interval staying in ON or OFF state, the adaptive PPSE considers past traffic load to determine whether to put the switch in ON or OFF state [7]. Figure 1 shows a finite state machine of adaptive PPSE mechanism. If pktCount, which denotes the number of packets received while ON state, exceeds threshold, the switch does not enter OFF state, that is, it keeps sending data without going to sleep. Therefore, average packet delay could be maintained under certain value even when there is high traffic load. If the pktCount is under threshold value, it indicates that traffic load is relatively low. In this case, the state of switch would enter OFF state, so that it can save power consumption.

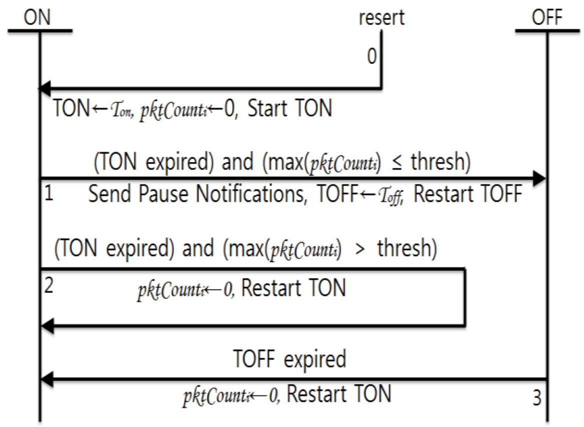

Figure 1. FSM of Adaptive PPSE mechanism

3. Enhanced Synchronizing Packet Coalescing Mechanism

In this section, the problems existing in adaptive PPSE mechanism are examined. To solve the problems, Enhanced Synchronizing Packet Coalescing (ESPC) mechanism is proposed to improve energy efficiency of small Ethernet switch.

Examining the simulation results obtained by an earlier study [7], there exist problems and constraints as follows.

First, Adaptive PPSE changes the state, to ON or OFF adaptively based on characteristic of incoming traffic. Thus, it achieves energy efficiency without increasing average packet delay significantly, even in case where traffic load is high or burst traffic happens suddenly in a short period. On the other hand, state transition takes place depending on the state of one specific port that has the highest traffic load among the ports existing in switch. The problem posed behind this fact is that adaptive PPSE cannot consider overall characteristic of network traffic. More specifically, unreasonable situation happens when only one link has pktCount exceeding threshold. Even though all the links except for one show very low value of pktCount, all the links in the switch stays in ON state. Another issue happens when substantial number of links have the pktCount value right below threshold. In this case, even if all the links have pktCount value under threshold, transition to OFF state can cause severe transmission delay.

Second, in case of applying fixed duty cycle, energy consumption rate increases significantly when incoming traffic load gets too high to fit in duty cycle. In particular, it gets worse when threshold value is set to low, for example, 1000 bytes, because the number of packet tends to exceed this threshold value in most time that switch spends in ON states. In other words, the switch mostly stays in ON state without sleeping, which means that energy consumption rate reaches to 100%. In this case, average packet delay drops significantly. However, energy saving cannot be expected at all. With use of high threshold value, for example, 5000 bytes, energy efficiency is growing gradually when traffic load exceeds duty cycle. As a result, the time that switch spends in ON state increases. In spite of longer ON state duration, average packet delay tends to increase a little bit. Ideal case is expected to show a linear increase of energy consumption rate, as traffic load grows.

In access network, average traffic load of switch is below 10% of link capacity. It is relatively low compared to core network. Its traffic flow varies significantly and it tends to have bursty pattern [8]. In terms of traffic load, much higher load than average is observed and much lower load is also observed in certain interval. Therefore, energy efficiency needs to be ensured maintaining average packet delay to be low level for each case.

Third, it should be reconsidered that applying same threshold to all the links is an appropriate approach. Ethernet switch is the equipment that is deployed in access network. For downstream, it lets several small-scaled terminals(from four to eight) connect to network. For upstream, it lets them connect towards internet core network. Most of internet services are implemented based on client-server model. Thus, traffic capacity of downstream and upstream is asymmetric. In general, traffic capacity of downstream is much bigger than that of upstream. In this sense, it is reasonable to set threshold value for down link and up link differently.

To alleviate the problems of adaptive PPSE as mentioned above, this study proposes enhanced synchronizing packet coalescing (ESPC) mechanism by means of variable threshold value. In ESPC mechanism, incoming traffic amount is measured during certain period. Based on the measured data, traffic characteristic is predicted for a future period. Finally, it adjusts threshold value to be suitable for the traffic head. In addition to state variables defined in adaptive PPSE, the proposed mechanism maintains additional variables for deciding traffic characteristic as follows.

- upPktCount[i]: An array to correspond to the number of up-links(i). It contains an amount of incoming traffic for the corresponding up-link during ON period.
- downPktCount: An amount of incoming traffic towards down-link during ON period.
- totalCount: A total amount of traffic incoming through all the links during ON period.
- threshold: A reference value which is compared to an amount of incoming traffic during ON state to determine if transition to OFF state should be made.
- Duty cycle (D): Percentage of time the switch spends in ON state. The value is defined as equation 1. By default, D is set to 10%, while TOFF is set to 100msec.
- longTermRate: Long-term average transmission rate of incoming traffic, which means an average transmission rate during period of (k * TON). The value of k is chosen to be optimal from the simulation result.
- shortTermRate: Short-term average transmission rate of incoming traffic, which means an average transmission rate during TON.

Once a packet is received during ON state, the packet is delivered to destination port. After that, traffic count (upPktCount or downPktCount) and total traffic count(totalCount) for the link receiving the packet is incremented correspondingly.

In next step, short-term and long-term average transmission rate of incoming packet are calculated using sliding window (TSW) [9]. TSW maintains three kinds of variables such as Win_length, Avg_rate, and T_front. Win_length is a parameter to keep history of previously arrived traffic. Depending on its value, traffic measurement can be done differently. Win_length with low value means that traffic is measured in short term. Hence, rapidly changing traffic can be measured precisely with low Win_length. However it is not suitable for measuring TCP traffic that tends to have relatively long-term connection while transmitting data. With use of high value for Win_length, average transmission rate can be measured relatively precisely for the long-term transmission, whereas sudden rapid change may not be measured properly. Thus, the proposed mechanism sets two windows, one of which for short-term measurement and the other for long-term measurement. Measuring average transmission rates with two separate windows, identifying the relationship between two results is helpful to measure the traffic more precisely. Based on the relation of shortTermRate, longTermRate and current threshold value, threshold can be adjusted. In this mechanism, threshold is adjusted as follows. It is incremented by α percent of transmission speed predicted using shortTermRate. Subsequently, threshold value is set to be the number of corresponding transmission packets. Optimal value of α is obtained from performance analysis. Additionally, longTermRate is used for deciding new duty cycle value depending on its relation with duty cycle transmission rate(D * link transmission rate). However, this paper does not deal with this point.

When timer expires in the ON period, the proposed mechanism determines whether the switch goes to OFF state or not depending on the relation between maximum value of upPktCount[i], downPktCount, and threshold. If maximum value is same as or more than threshold, state variables are set to make the state constantly stay in ON state without transition. Otherwise, state variables are set to make state transit to OFF state. In the case of OFF state, data transmission is stopped and all the ports lose the power. If timer expires in OFF state, state variables are initialized and then the state goes to ON. From now on, the procedures as mentioned above would be performed repeatedly.

4. Performance Evaluation

4.1. Simulation Environment

Performance of the proposed mechanism was evaluated by simulation. CSIM 20 simulator was used as simulation tool [10]. Network architecture was developed in a same form as that from previous study [7]. The network traffic is modelled with a concept that all the traffic incoming to the switch through all links are aggregated into one link. Traffic transmission rate is determined by packet size of an aggregated link. Once traffic transmission rate is obtained, state transition would be determined. Script file for simulation was coded to run its procedures as mentioned above. Table 1 describes parameters that are required to evaluate performance.

Table 1. Simulation parameters

parameters	value
Ethernet Link capacity	10 Gbps
Duty Cycle	10 %
Toff	100 msec
Ton	11.11 msec
α	10 %
Win_length	11.1 msec

In terms of incoming traffic pattern, we consider two different cases. In first case, it is assumed that traffic pattern follows Poisson distribution with 1500 bytes of fixed packets. Script file was implemented to determine an average arrival rate depending on predefined traffic load. In second case, it is assumed that traffic has bursty nature. To evaluate performance under bursty condition, traffic trace file was used, which was generated by synthetic traffic generator [11].

Performance was evaluated by simulating with three different threshold (1000 fixed, 5000 fixed and variable value determined by predicted traffic). Traffic load was changed ranging from 1% to 30%. In order to analyze performance, comparative experiment was carried out by measuring energy consumption rate and average packet delay.

Energy consumption rate can be produced using equation 2. Another assumption related to energy consumption, is that energy consumption rate is 100%in ON state and 10% in OFF state. Aon and Aoff denote a time duration that switch spends in ON state and OFF state respectively.

$$Pa = 100 * (Aon + 0.1*Aoff)/(Aon + Aoff) \ (\%) \quad (2)$$

Packet delay indicates time duration spent delivering the packet to destination link since the packet arrives the switch. In other words, it means spending time from the point when packet arrives in network interface card to the point when packet is delivered to destination link through the switch.

4.2. Simulation Results Analysis for Poisson traffic

Simulation results for traffic with Poisson distribution is shown in Figure 2. One graph presents ideal case. The other three graphs illustrate energy consumption rate for different threshold value. Traditional switch shows 100% of energy consumption rate because it always stays in ON state independent on traffic load. In energy efficient Ethernet environment, an ideal case means that energy consumption rate increases linearly as traffic load grows.

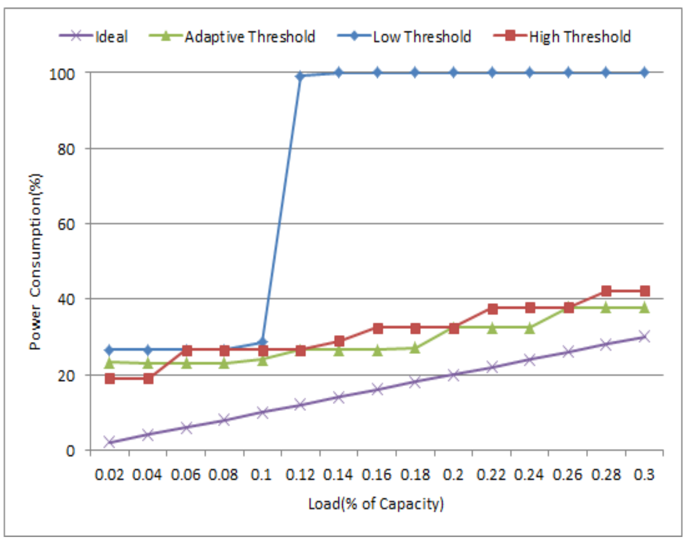

Figure 2. Power Consumption (Poisson Traffic)

In case of low threshold and high threshold, performance evaluation results showed very similar curve with previous work. In contrast, energy consumption rate is observed higher than previous work. It is because our assumption is different from previous work in energy consumption of OFF state. This study assumes 10% energy consumption happens in OFF state whereas previous study assumes no energy consumption. With adaptive mechanism proposed in this study, the results indicate that energy consumption rate is lower than that of fixed threshold approach in most cases with different traffic loads. In particular, ranging 6-10% of average traffic load of access network, proposed mechanism shows 23% of energy consumption rate, while traditional mechanism records 26.5%. There are significant improvement in energy efficiency.

Figure 3 illustrates simulation results for average packet delay. In case of low threshold, average delay drops at the lowest level. In particular, in case where traffic load is over 12% of link capacity, link is operated on ON state all the time, which means average delay is almost zero. In case below 5%, average packet delay of proposed mechanism was recorded 44.8msec, which is lower than that of high threshold case (45.8 msec). Average traffic load of access network is 6-10%. In this range, adaptive threshold mechanism shows higher average packet delay than fixed threshold case. Nevertheless, the difference is not significant. In case over 15%, average delay tends to grow gradually. However, since total packet delay is below 50msec, the delay does not affect service quality even for real-time application.

With the proposed algorithm, average packet delay increases by tiny amount, compared to traditional algorithm. Nevertheless, it improves overall performance because it reduces energy consumption rate significantly.

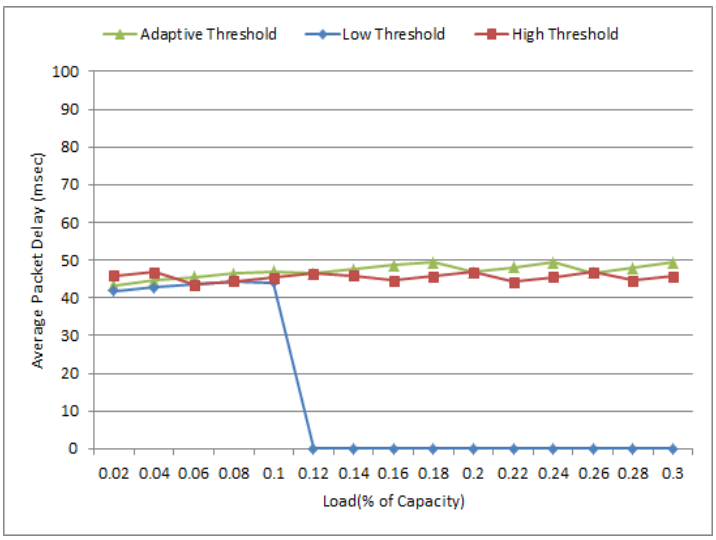

Figure 3. Average Packet Delay (Poisson Traffic)

4.3. Simulation Results Analysis for Bursy traffic

To generate burst traffic trace, we set synthetic packet generator parameter as previous work [7] did. Figure 4 illustrates simulation results for burst traffic. Likewise Poisson traffic, performance evaluation results of adpative PPSE show similar curve with previous study.

With the proposed mechanism, energy consumption rate records lower than that of fixed threshold case in most traffic loads. In particular, the proposed mechanism shows 22.7% of energy consumption rate, while traditional mechanism shows 27.1% in the range of 6-10%, which is average load. This result indicates that proposed mechanism contributes to energy efficiency.

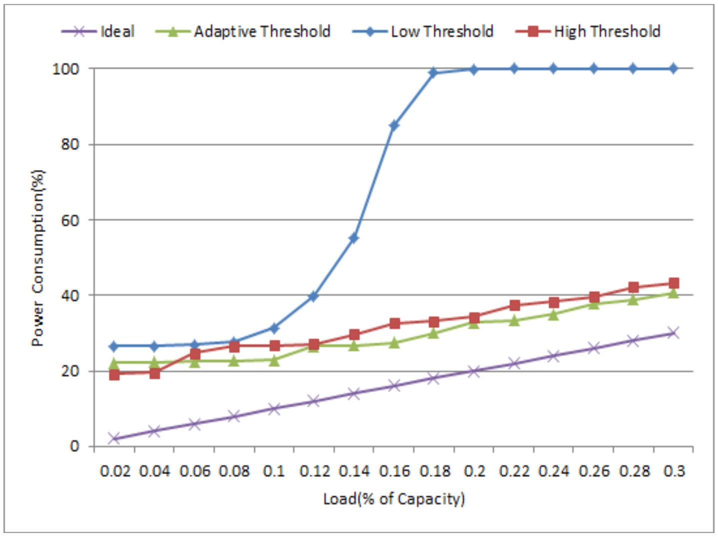

Figure 4. Power Consumption (Burst Traffic)

Figure 5 illustrates simulation results for average packet delay. With low threshold, average packet delay drops at the lowest level. In particular, in case where traffic load is over 18% of link capacity, average delay stays almost zero because link is operated in ON state all the time. In case below 5%, the proposed mechanism shows 45.2 msec of average packet delay, which is lower than that of high threshold (47.2 msec). In the rage of 6-10%, which is average traffic load of access network, average packet delay was higher than fixed threshold. However, the difference was insignificant. Since total packet delay is 48.5msec, the delay does not degrade service quality of real-time application.

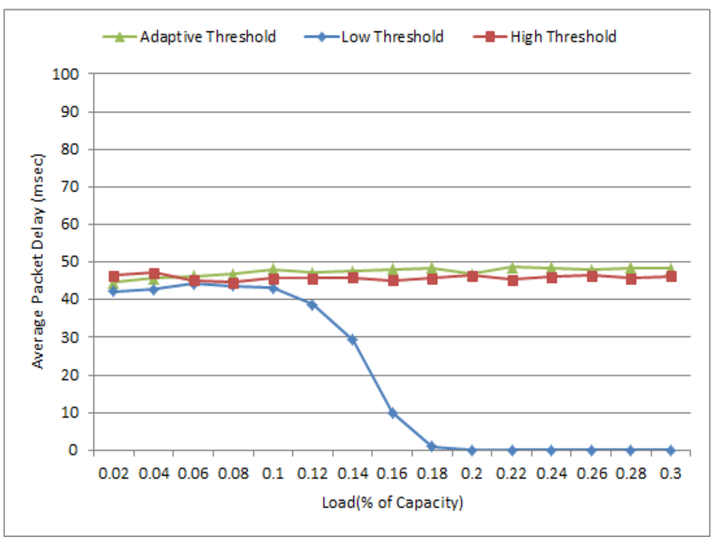

Figure 5. Average Packet Delay (Burst Traffic)

5. Conclusion

Due to sharp rise of real-time multimedia traffic based on Internet, network equipments are required to have high speed data transmission capability. At the same time, energy consumption also grows fast. Severe energy consumption becomes one of main challenges in this area. To overcome the problem, research efforts need to be focused on developing energy-efficient communication protocol for access network equipments such as Ethernet switch, network interface card as well as core router. In this sense, this paper proposes Enhanced Synchronizing Packet Coalescing mechanism that is able to adaptively adjust threshold value based on incoming traffic characteristic.

In this mechanism, an amount of incoming traffic is measured in certain period. Based on the measured value, it predicts traffic characteristic of coming period. Finally, it adjusts threshold value to be optimized corresponding to traffic load, which results to improving energy efficiency.

Performance evaluation results by simulation indicate that proposed algorithm outperforms traditional algorithm at the expense of acceptably excessive packet delay. Specifically speaking, the proposed algorithm reduced energy consumption rate significantly, although average packet delay increases little bit in the proposed algorithm compared to traditional one. Currently, input traffic model is being simulated considering internet traffic characteristic. In the future, this mechanism needs to be more investigated with focus on its effect on TCP congestion control mechanism.

Acknowledgements

This work (Grants No.C0029530) was supported by Business for Cooperative R&D between Industry, Academy, and Research Institute funded Korea Small and Medium Business Administration in 2012.

References

[1] M. Gupta and S. Singh, "Greening of the Internet", in Proc. ACM SIGCOMM, **(2003)** August, pp. 19–26.
[2] K. W. Roth, F. Goldstein and J. Kleinman, "Energy consumption by office and Telecommunications equipment in commercial buildings", vol. 1: energy consumption baseline, National Technical Energy Consumption Service (NTIS), U. S. Department of Commerce, **(2002).**
[3] C. Gunaratne, K. Christensen, B. Nordman and S. Suen, "Reducing the energy consumption in Ethernet with Adaptive Link Rate (ALR)", IEEE Trans. Computers, vol. 57, **(2008)**, pp. 448–461.
[4] K. Christensen, P. Reviriego, *et al.*, "IEEE 802.az: the road to energy efficient Ethernet", Communications Magazine, **(2010).**
[5] IEEE P802.3az, Energy Efficient Ethernet Task Force, **(2010).**
[6] P. Reviriego, J. Hernandez, D. Larrabeiti and J. Maestro, "Performance evaluation of Energy Efficient Ethernet", IEEE Communication Letters, **(2009).**
[7] M. Mostowfi and K. Christensen, "Saving energy in LAN switches: New methods of packet coalescing for Energy Efficient Ethernet", Green Computing Conference and Workshops(IGCC), **(2011),** pp. 1-8.
[8] A. Odlyzko, "Data Networks are Lightly Utilized, and Will Stay That Way", Review of Network Economics, vol. 2, **(2003)**, pp. 210-237.
[9] D. Clark and W. Fang, "Explicit Allocation of Best Effort Packet Delivery Service", ACM Transaction on Networking, **(1998).**
[10] H. Schwetman, "CSIM19: A Powerful Tool for Building System Models", Proceedings of the 33rd Winter Simulation Conference, **(2001)** December.
[11] R. Jain, "The Art of Computer Systems Performance Analysis: Techniques for Experimental Design", Measurement, Simulation, and Modeling, Wiley-Interscience, New York, **(1991)**.

Authors

Young-Hyun Kim

Young-Hyun Kim received the B.S. degree computer science from Chonnam National University, Korea, in 1985, M.S. degree in computer science from Chosun University, Korea, in 1993, respectively. He joined in Computer Center, Gwangju National University of Education, Korea, since 1989, where he is currently a Section Chief. Recently, he has participated in establishment of campus network. His current interests include Energy efficient ethernet, Network, Wireless sensor network, DB.

Sung-Keun Lee (Corresponding Author)

Sung-Keun Lee received the B.S., M.S., and Ph. D. degrees in electronics engineering from Korea University, Seoul, Korea, in 1985, 1987, and 1995, respectively. From 1987 to 1992, he was with Samsung electronics Co., Ltd, Korea. He joined the department of Multimedia Engineering, Sunchon National University, Suncheon, Korea, in 1997, where he is currently a Professor. His research interests include energy efficient Ethernet, wireless sensor network, multimedia communication and Internet QoS.

Jin-Gwang Koh

Jin-Gwang Koh received the B.S., M.S., and Ph. D. degrees in Computer Science and Engineering from Hongik University, Seoul, Korea, in 1982, 1984, and 1997, respectively. He was also Visiting Scientist to Oregon State University for the year 1997-1998. He was vice president of Korean Institute of Information Scientists and Engineers. He joined the department of Computer Science and Engineering, Sunchon National University, Sunchon, Korea, in 1988, where he is currently a Professor. His research interests include Database, wireless sensor networks.