Design and Analysis of Ethernet Aggregation to XGMII Framing Procedure

Youjin Kim, Jaedoo Huh,

Electronics and Telecommunication Research Institute (ETRI)
161 Gajong-Dong, Yusong-Gu, Taejon, 305-350, Korea
{youjin, jdhuh}@etri.re.kr

Abstract - This paper suggests the Ethernet aggregation to XGMII framing procedure (EAXFP) mechanism to economically combine the traffic adaptation technology with the link aggregation method in designing 10 Gigabit Ethernet (10 GbE) interfaces. This design sidesteps the data-loss issues that can result from designing an interface with only one link. The most critical issue in relation to the link aggregation interface is the algorithm used to control frame distribution between the ten ports. The proposed EAXFP mechanism offers an efficient link aggregation method as well as an efficient frame distribution algorithm, which maximize the throughout of the 10 GbE interface. In the experiment and analysis of the proposed mechanism, it was also discovered that the 10 GbE interface that uses the proposed EAXFP mechanism significantly reduced the packet loss rate. When there will be heavy traffic loads come about in the future, the proposed EAXFP mechanism assures an efficient and economical transmission performance on the router system.

Keywords: router, 10GE, Ethernet, GFP

1 Introduction

In this paper, efficient models or protocols in designing the 10 GbE interface was studied based on the requirements of the NGN. These models and protocols were confirmed through a comparison of edge router systems performing with the 10 GbE. Below are two suggestions for designing the proposed router system in a stable and effective way. First, the effective design of the 10 GbE interface is necessary because with the NGN, the uplink from the medium-capacity edge router to the large-capacity backbone router is supposed to be connected to the MAN and WAN based on the 10 GbE. Various technologies, such as the Generic framing procedure (GFP), are also necessary in order to effectively forward Ethernet traffic to the SONET/SDH networks on a large scale [1, 9]. Based on this, the Ethernet aggregation on the XGMII framing procedure (EAXFP) mechanism is proposed to organize the ten units of 1 GbE with three units of the four-gigabit capacity NPs instead of using a very expensive one 10-gigabit-capacity NP [2, 7]. Second, traffic adaptation technology such as the GFP is necessary for several units of NP to interface through the proposed EAXFP mechanism rather than the simple link aggregation of multiple ports mechanism. Combination of the traffic adaptation technology with the link aggregation method can help design the EAXFP framer.

This paper is organized as follows. Section 2 includes the case study that explains a generic framing procedure (GFP) as traffic adaptation mechanism. The proposed EAXFP mechanism is described in section 3. The experiments on an implemented EAXFP mechanism are presented in section 4. Finally, conclusion and future works are given in the last section.

2 Generic framing procedure (GFP)

The Generic framing procedure (GFP) is a recently standardized traffic adaptation protocol for broadband transport applications that has emerged in response to this demand [8, 9]. It provides an efficient and QoS-friendly mechanism to map either a physical layer or logical link layer signal to a byte-synchronous channel. However, as the process of standardization for 10 GbE was not easy, the development of a new LAN technology, which provides high speeds of over 10 Gbps, will take longer unless significantly more efficient forwarding techniques emerge. Consequently, applications that require speeds of over 10 Gigabit will need to combine the traffic adaptation technology with the link aggregation method. Such technologies effectively offer high-speed

transmission through multiple low-speed links with previously developed forwarding technologies.

We survey the example of 'Ethernet over SONET/SDH using GFP' illustrated in Fig. 1 [9]. Widespread acceptance of the Ethernet and the emergence of 1 GbE and 10 GbE have generated interest in transporting Ethernet frames across SONET/SDH networks. One such approach, ITU-T X.86 Ethernet over SONET/SDH (EoS), relies on familiar HDLC technology. Rather than encapsulate IP/PPP packets, EoS encapsulates complete Ethernet frames in HDLC packets [5]. The GFP mechanism for existing Ethernet over SONET/SDH consists of combining high-speed, circuit-oriented physical ports with many different low-speed, packet-oriented physical ports. Another method known as the link aggregation technology combines many low-speed physical ports into a single logical port, which represents all the capacities of each physical port.

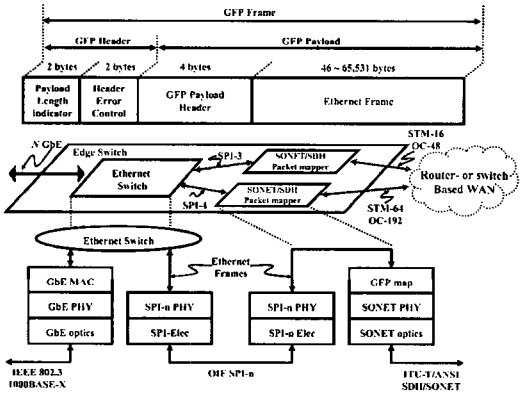

Fig. 1. Ethernet and GFP frame relationships

There are many advantages of the link aggregation (LA) technology. First of all, it enables the network to reuse communication devices with low capacities per port when users need circuits with higher capacity per port for their communication devices without extra cost. In addition, the calculation overheads on a single port can be less than that for various ports with regard to the routing protocol. Due to the efficiency of the link aggregation technology, it is widely used in many communication devices. For example, the LA function is integrated in most new gigabit Ethernet devices as well as network processors. Although this technology ties many low-speed interface lines to a single logical high-speed interface by linking them together between two devices, it is less utilized in terms of efficiency. Therefore, the traffic adaptation mechanism is required to link all low-speed lines into one high-speed line. There is a method developed previously which implements only the traffic adaptation mechanism. Unfortunately, communication devices in this case cannot recognize one high-speed logical interface but instead see many low-speed interfaces. This paper will propose the EAXFP mechanism described in section 3, which combines the existing traffic adaptation mechanism and the link aggregation technology together. Since network processors are commonly used, the proposed EAXFP mechanism offers link aggregation to create high-speed interfaces through many low-speed NPs. In other words, it offers XGMII interface by combining 10 units of a 1 GbE port into three units of 4-gigabit capacity NPs, instead of using a single unit of very expensive 10 Gigabit capacities NP. In the EAXFP mechanism, ten 1 GbE ports are aggregated to a single XGMII, while eight 1 GbE ports are aggregated to a single SPI-4.2 in the GFP design for the Ethernet over SONET/SDH. It also provides an effective link aggregation methodology by implementing the EAXFP mechanism.

3 Design of the proposed EAXFP framer

The proposed 10 GbE line card, illustrated in Fig. 2, integrates the following functions: the physical layer block of XGXS and; the EAXFP framer block, which is a combination of 10 MAC interfaces of Gigabit Ethernet; the network processor (NP) block, which processes the data packet; and the line card processor block, which undertakes IPC communication with the routing processor (RP): three NPs are connected to the line card processor (LP) via the PCI bus [4]. Each NP is connected to the switch fabric and has four GMIIs.

By using three NPs, 10 GMIIs are combined and connected to the EAXFP framer built by using a FPGA. The EAXFP framer is connected to the 10 GbE physical layer of XGXS. The EAXFP framer is responsible for multiplexing incoming frames from 3 NPs and provides one 10 GbE to the physical layer by aggregating ten 1 GbEs. In addition, by connecting the XGMII interface to the XPHY, incoming frames from the XPHY are de-multiplexed in the EAXFP framer.

The NP extracts the Gigabit Ethernet frame from received data through GMII and process information such as packet classification, address lookup, forwarding, and traffic management for layer 2/3/4+ switching. It is also responsible for sending and receiving data packets to the switch fabric through the high-speed switch link. The NP supports a four-channel GMII interface, a two-channel high-speed switch link, and a one-channel PCI interface. The LP is based on the high-performance microprocessor interfaces between the NP and the PCI bus.

The EAXFP framer requires the following functions which are similar to the link aggregation (LA) methods of IEEE 802.3ad standard: (1) the frame distribution function which handles the order arrangement between frames at the time of transmission; (2) the way in which MAC addresses should be addressed; and (3) the solution at the time of link change in the process of the frame distribution.

These three requirements refer to the case where MAC clients consider the LA group to be a link by defining more than one link as the LA group. When the EAXFP design is linked with the physical interface of one 10 GbE the link aggregation methods differ from those specified in the IEEE 802.3ad. Therefore, what should be considered most important in the proposed design is the order arrangement function between frames at the time of transmission and reception. Wrong order arrangement between frames significantly reduces network efficiency especially in the case of streaming-oriented data.

The processing load of the 10 GbE must be fairly distributed between the three Network processors. To accomplish this there must be flow control even during the mutual conversion between GMII and XGMII. These requirements can be fulfilled by controlling the hardware of the EAXFP framer, and through the control of the software that manages the hardware of the EAXFP framer.

The EAXFP framer contains four blocks: (1) the frame multiplexer block, which aggregates ten 1 GbE frames to one 10 GbE frame; (2) the frame inverse multiplexer block, which converts one 10 GbE frame into ten 1 GbE frames; (3) the processor interface block; and (4) the loop back function block as shown in Fig. 3.

Frame order discord and frame duplication between the ten GMIIs and the one XGMII should not take place if the 10 GbE frames are intended to be transmitted using EAXFP framer. For an efficient frame distribution, two processes are proposed.

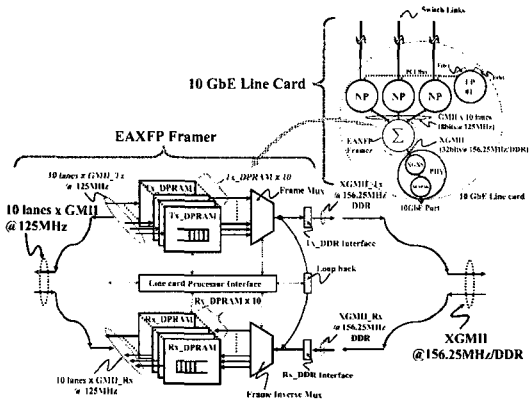

Fig. 2. The configuration of the proposed 10 GbE line card and EAXFP framer block diagram

The first process is the frame composition method, which supports the maximum size of 1 GbE using the padding. The next process is the frame order arrangement and distribution using the round-robin scheduler. These two processes are referred to as "the padding added round robin (PARR) frame distribution algorithm." In the EAXFP framer, as shown in Fig. 2, the maximum 1 GbE frame is constructed in the Tx_DPRAM. Frames that are shorter than the maximum 1 GbE frame are padded with the idle value (0x000001112).

If this method is used, frame order discord does not take place, and so the data packet can be transmitted to any of the link units and there will be no need to fix the links using the MAC address on the reception side. The implementation based on the PARR is simpler in comparison to existing static and adaptive frame distribution methods. This method, nevertheless, can be inefficient for relatively short frame lengths, since all transmitted frames are padded to the maximum frame length. How-ever, since the EAXFP framer is implemented with hardware logic for real-time pad-cing, there is no significant drop in efficiency.

As in Fig. 3, the second process makes a 10 GbE frame by sequentially sending packets to the frames constructed by the first process through the round-robin scheduler. Since the maximum length of the frames constructed using padding in the first process is fixed, it is possible to use the round-robin scheduler. This operation is carried out in the Tx_DPRAM and Frame Mux as shown in Fig. 3.

4 Performance Analysis and Experiment

The round robin scheduler of the EAXFP serves backlogged sessions sequentially in a fixed order. When the scheduler finishes transmitting a packet from session 1, it moves on to session 2 and checks to see if there are any packets waiting for to be transmitted. After the scheduler has finished examining all of the sessions, it returns to the first session.

We analyze the occupancy of each queue in the round robin scheduler of EAXFP using the following assumptions: a maximum of packets could arrive during one round into queue i; the probability of departure from the queue i is C_i = 1 [6]. The parameters from the design of the EAXFP framer are applied to the round robin scheduler in whom all sessions are identical with the following values as defined in Table 1. Repeating the same procedure for all other queues, we would then be able to find the performance of the round robin scheduler of EAXFP framer based on reference [6].

The wait time in units of seconds is simply defined as

$$W = \frac{Q_a T}{(1 - s_0) A_i} \qquad (1)$$

The performance evaluation of the 10 GbE line card by using the suggested EAXFP mechanism is described as follows: The ten 1 GbE ports and one 10 GbE port in the suggested edge router are connected to the measuring instrument. For the interoperability test, one 10 GbE port is connected to the 10 GbE port in another router, and each 1 GbE port is connected to the commonly used router equipment [2].

The measurement result of throughput to ten 1 GbE ports and one 10 GbE port in the same way is the same as in Fig. 6. The throughput value being measured by outputting packets through the 10 GbE line card with the EAXFP framer at the time of putting traffic simultaneously to ten 1 GbE ports is shown in Fig. 6. This proves that EAXFP framer that organizes the ten units of 1 GbE with three units of the four-gigabit capacity NPs instead of using a one ten-gigabit capacity NP to be a normal process without packet loss

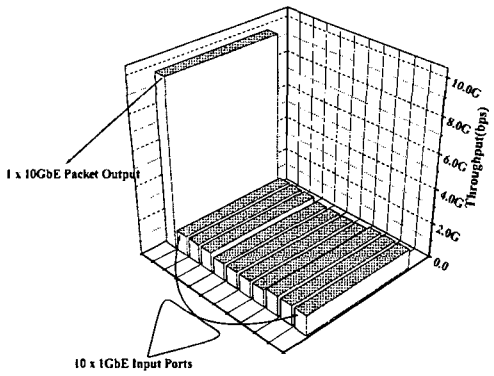

Fig. 6. Throughput at the 10 GbE using EAXFP framer

5 Conclusions and Future Works

The 10 GbE technology is expected to replace existing SONET/SDH in metropolitan area networks. Since 10 GbE is recognized as the most economical technology from which to construct metropolitan and wide area networks without using the optical forwarding devices of SONET/SDH, Internet service providers expect that the 10 GbE will break down the distinctions that have separated local area networks from metropolitan and wide area networks.

To make use of this architecture it would be necessary to design 10 GbE ports efficiently, as the edge routers' uplink to backbone routers would be normally configured as MAN and WAN through 10 GbE in the NGN. In addition, technologies like the Generic framing procedure (GFP) would be required to transmit Ethernet traffic efficiently over the full-blown SONET/SDH networks. In order to meet such requirements, this paper proposes the EAXFP mechanism, which implements three 4 Gbps-class NPs to aggregate ten 1 Gbps ports rather than constructing one 10 GbE interface by using a high-cost 10 Gbps-class NP. The proposed EAXFP mechanism provides an efficient link aggregation method and frame distribution algorithm to optimize the 10 GbE interface operation. Furthermore, the proposed padding added round robin (PARR) frame distribution algorithm creates a maximum frame size up to 1 GbE, and arranges the frames by using the round robin scheduler.

It is expected that the proposed EAXFP mechanism would be a valuable resource for designing the 10 GbE interfaces in router systems. It might further provide valuable resources for designing the 40 GbE interface based router systems for the NGN.

For future works, the proposed EAXFP mechanism will be applied to the design and implementation of 40 GbE possible line card for the NGN, will provide various amendments to the PARR frame distribution algorithm.

References

1. D. Cavendish et al., "New transport services for next-generation SONET/SDH systems," IEEE Communications Magazine, vol. 40, no. 5, pp. 80~87, May. 2002.
2. D. Kim, S. Lee, and C. Choi, "Trends of 10 Gigabit Ethernet switch development in Korea," IEEE Pacific Rim Conference on Communications, Computers and signal Processing, vol. 2, pp. 1032~1035, Aug. 2003.
3. S. Lee, B. Joo, and H. Jung, "Implementation and performance analysis of high speed packet switching system with gigabit interface," Proceedings of the Institute of Electronics Engineers of Korea Conference, Nov. 2002.
4. Y. Kim, H. Jung, and K. Cho, "Design and evaluation of redundant IPC network adequate for an edge router," Lecture Note in Computer Science, no. 3126, pp. 279~290, Aug. 2004.
5. ITU-T, "Generic framing procedure (GFP)," Rec. G.7041/Y.1303, 2001.
6. G. Fayez, "Computer communication networks: analysis and design," North Star Digital Design, Inc. Victoria, Canada, 2001.
7. C. Choi, B. Joo, and D. Kim, "Design and implementation of 10 Gigabit Ethernet frame multiplexer/de-multiplexer," Proceedings of the Institute of Electronics Engineers of Korea Conference, vol. 1, pp. 378~381, Jul. 2003.
8. E. Hernandez-valencia, M. Scholten, and Z. Zhenyu, "The generic framing procedure (GFP): an overview," IEEE Communications Magazine, vol. 40, no. 5, pp. 63~71, May. 2002.
9. P. Bonenfant and A. Rodriguez-moral, "Generic framing procedure (GFP): the catalyst for efficient data over transport," IEEE Communications Magazine, vol. 40, no. 5, pp. 72~79, May. 2002.