Flush Optimizations to Guarantee Less Transient Traffic in Ethernet Ring Protection

Kwang-Koog Lee and Jeong-dong Ryoo

Ethernet ring protection (ERP) technology, which is defined in ITU-T Recommendation G.8032, has been developed to provide carrier grade recovery for Ethernet ring networks. However, the filtering database (FDB) flush method adopted in the current ERP standard has the drawback of introducing a large amount of transient traffic overshoot caused by flooded Ethernet frames right after protection switching. This traffic overshooting is especially critical when a ring provides services to a large number of clients. According to our experimental results, the traditional FDB flush requires a link capacity about sixteen times greater than the steady state traffic bandwidth. This paper introduces four flush optimization schemes to resolve this issue and investigates how the proposed schemes deal with the transient traffic overshoot on a multi-ring network under failure conditions. With a network simulator, we evaluate the performance of the proposed schemes and compare them to the conventional FDB flush scheme. Among the proposed methods, the extended FDB advertisement method shows the fastest and most stable protection switching performance.

Keywords: Carrier Ethernet, Ethernet ring protection, MAC address learning, filtering database flush, multi-ring.

I. Introduction

Currently installed metropolitan area networks (MANs) were normally built upon ring topologies supported by synchronous optical network (SONET)/synchronous digital hierarchy (SDH). As SONET/SDH technology is based on time-division multiplexing (TDM), it is far more suitable for voice-centric traffic. However, with the rise of the Internet and the expansion of broadband worldwide, the majority of the services that are provided across MANs are data traffic. Most data traffic originates and terminates at Ethernet local area networks (LANs). Ethernet is evolving to be a dominant technology for carrier-grade networks due to its simplicity, cost effectiveness, and flexibility. It is rapidly gaining importance as a prevailing solution to support the transport of future Internet protocol (IP) services [1].

As Ethernet technology continues to make considerable progress toward the transport network, it has been challenged by service providers that need fast protection switching capability to guarantee carrier-grade availability. Traditionally, Ethernet focusing on the scope of LAN relies on a spanning-tree protocol (STP) to build an active topology for data frame forwarding while ensuring loop avoidance [2]. The STP creates a logical topology in the form of a spanning tree where the path from the root node to every other node is the shortest-path with minimum hop-count. However, the efficiency of the formed logical topology depends on a choice of the root node. Furthermore, the STP approach acts slowly upon any topology change caused by a failure or the recovery of a failure within the network. Even with advanced STP-based approaches, such as the RSTP and the MSTP [3], [4], the convergence time after a topology change is still too long to meet the sub-50 ms protection requirement for carrier-grade Ethernet networks. For

Manuscript received Sept. 15, 2009; revised Jan. 4, 2010; accepted Jan. 18, 2010.

Kwang-Koog Lee (phone: +82 42 860 6723, email: kwangkoog@etri.re.kr) and Jeong-dong Ryoo (corresponding author, email: ryoo@etri.re.kr) are with the Internet Research Laboratory, ETRI, Daejeon, Rep. of Korea, and also with the Department of Broadband Network Technology, University of Science and Technology, Daejeon, Rep. of Korea.

doi:10.4218/etrij.10.1409.0097

this reason, there have been several efforts in which the main goal is to provide better reliability and resilience in a carrier-grade Ethernet network. Such efforts led to the appearance of ring-based protection approaches such as resilient packet ring (RPR) [5], [6] and Ethernet ring protection (ERP) [7].

RPR, defined in IEEE 802.17, is an MAN technology supporting data transfer among stations interconnected in a dual ring configuration. It provides not only a high resilience, which allows fault protection switching within 50 ms, but also management of excess information rate (EIR) traffic under traffic congestion and protection scenarios. However, it introduces a new media access control (MAC) header, whose format is different from that of typical Ethernet, and a new set of complex protocols and algorithms in terms of topology discovery and fairness. These drawbacks increase development and deployment costs and weaken RPR's economic viability.

ERP, defined by G.8032, has been developed on a principle of utilizing generic mechanisms inherited from the traditional Ethernet MAC and bridge functions. It is designed for ring topologies and developed as a standardized alternative to replace the STP to change the port status without requiring complex computation, provisioning overhead, or excessive information exchange, so as to achieve fast sub-50 ms protection switching in a simple way. Consequently, ERP provides rapid service restoration that delivers SONET/SDH-grade resilience at Ethernet cost for small-to-medium applications.

ERP works on the basis of filtering database (FDB) flush. On protection switching for a failure or a failure recovery, all ring nodes remove all learned MAC addresses in their FDBs for a changed ring topology. Then, each ring node has to broadcast unknown data frames until source MAC learning is completed. The duplicated frames, however, make a ring network suffer from a large amount of traffic overshoot several times greater than the steady state traffic. When such flooding traffic volume is far greater than the link capacity, the majority of frames are lost or delayed due to queuing in a buffer. Since the traffic flooding will require more link capacity, a flush operation should be avoided as much as possible.

This paper introduces four FDB flush optimization schemes to provide less traffic overshoot, ring-centric FDB flush (*r-Flush*), port-based FDB flush (*p-Flush*), extended selective FDB advertisement (*e-ADV*), and *flush triggering*. The first scheme, r-Flush, makes each ring node perform an FDB flush operation on the ring ports only, excluding its client subnet port. The second scheme, p-Flush, uses the port information related to the receipt of a ring automatic protection switching (R-APS) protocol message, which is used to coordinate the protection switching activities among the nodes on the ring. It allows each node to flush the FDB entries associated with the reception port. The third scheme, e-ADV, is an enhancement of the previously proposed selective FDB advertisement scheme in [8]. The extended scheme introduces added functions by which the previous scheme can be operated in multi-ring networks. The last scheme, flush triggering, utilizes the optimality condition that only the nodes on the active path of an upper ring need to be flushed in the case of a topology change in a lower ring. This scheme can be combined with any of three proposed schemes.

This paper is organized as follows. Section II summarizes related works. Section III introduces the ERP technology and network stability issues in ERP. Then, in section IV, this paper shows how the proposed four methods are designed and used for ERP. Their protection switching performance is evaluated and discussed in section V. Finally, conclusions are drawn in section VI.

II. Related Works

Ryoo and others introduced the ERP technology to relate its importance for future packet transport networks [9]. This paper outlines the concepts in the ERP switching and discusses the fundamental operation principles on which the R-APS protocol works.

In [8], Lee and others proposed the selective FDB advertisement, which provides the protection mechanism for a single ring network. Since the method fully complies with the node architecture and R-APS protocol message relay model defined in G.8032, it guarantees an effective and realistic protection solution for a single Ethernet ring network.

Rhee and others proposed the FDB flipping method to solve the traffic overshoot problem [10]. When a signal fail (SF) occurs, this mechanism makes the nodes adjacent to the failed link (NAFs) send R-APS (SF, flip) messages to inform other nodes of the addresses that are impacted by protection switching. As nodes receiving these messages update their FDBs using the address information in the R-APS (SF, flip), the FDB flipping method features an immediate transition to the steady state. However, this method has two major drawbacks. First, the R-APS (SF, flip) message always has to be terminated and regenerated in a hop-by-hop manner. If there are a large number of flipped addresses resulting in multiple segmented R-APS (SF, flip) messages, it might lead to a prolonged protection switching because of a significant processing delay in each node. Secondly, it requires an impractical assumption that all ring nodes must have identical MAC addresses in their FDBs.

The network stability issues have also been discussed in RPR technology. In [11], Kvalbein and others analyzed both basic [5] and enhanced bridging algorithms [6] defined in the RPR standard through simulation. The basic bridging

algorithm leads to poor bandwidth utilization due to the loss of spatial reuse for transparent bridging to the RPR network. The enhanced algorithm effectively reduces the load imposed on the network without flooding, whereby it enables bridges to maintain global and local tables with respect to the addresses of both a bridge and a local host in the extended packet format.

Setthawong and Tanterdtid suggested a solution for bridging RPR networks that do not require the flooding of inter-ring packets [12]. In their solution, bridges periodically learn about all the nodes in the bridged network by using the attribute discovery (ATD) packet in the RPR and update the next-hop bridge to each of those nodes in a newly defined global topology table (GTD). This scheme is similar to our proposed e-ADV mechanism, but it is operated for only interconnected RPR rings, and the ATD packets in their scheme, including MAC addresses of nodes, have to be transmitted periodically regardless of the normal or protection state of a ring. Too many bridges in RPR rings can lead to inefficient bandwidth utilization and processing complexity.

Several other types of resilient Ethernet-based rings have been proposed. Zhong and others proposed a new optical resilient Ethernet ring (RER) for MANs [13]. They described in detail the basic RER system design issues including RER architecture, frame format, frame forwarding, protection, and interconnection methods for multiple rings.

III. Network Stability Issues in Ethernet Ring Protection Switching

1. Overview of Ethernet Ring Protection

The ERP technology defined by ITU-T recommendation G.8032 [7] has been developed to provide carrier-grade resiliency in Ethernet ring networks. To achieve fast protection switching, ERP uses one specified link called a ring protection link (RPL) to prevent traffic loops over the ring under normal conditions. It is unblocked under failure conditions in any place other than the RPL. Each end of the RPL is attached to two ring nodes, called RPL owner and RPL neighbor, which are responsible for blocking or unblocking their RPL port depending on ERP states. To coordinate protection activities, ERP defines an R-APS protocol, which is added on operations, administration, and maintenance functions of Ethernet [14]. The R-APS protocol messages are conveyed via a control channel called an R-APS channel, which is separated from the data channel for client traffic by a different VLAN ID. An Ethernet ring can be extended to interconnection of multiple rings. In such a case, a lower ring (subring) is connected by single or multiple shared links to an upper ring (major ring) through the use of interconnection nodes.

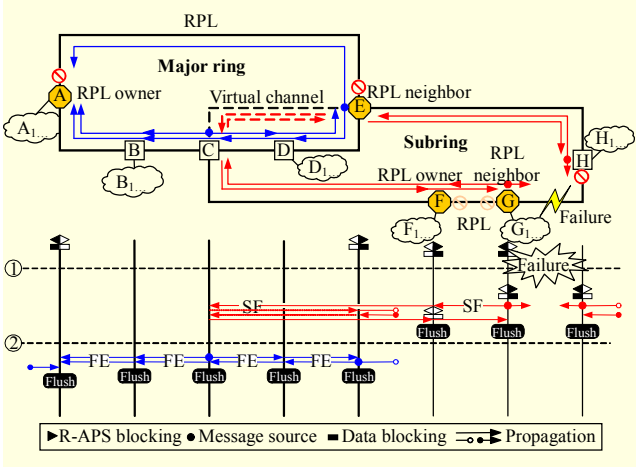

Fig. 1. Single link failure procedure on multi-ring network.

An example of the ERP mechanism under failure is illustrated in Fig. 1. The major ring is composed of nodes A through E and forms one complete circle. Meanwhile, the subring consists of nodes C through E and is in the shape of a horseshoe. Nodes C and E are the interconnection nodes for both rings. The RPL owner and RPL neighbor in each ring initially block their RPL ports for both the R-APS channel and data channel to create a logical loop-free topology. Additionally, an R-APS virtual channel between node C and node E provides R-APS connectivity of the subring. The R-APS virtual channel can be created with a different VLAN ID in the major ring space.

When link G-H fails, the NAFs, nodes G and H, detect the local SF and immediately block the port facing failure. Each NAF subsequently transmits an R-APS (SF) message in both directions and flushes its FDB to relearn clients' locations on the changed ring topology. Upon receipt of the R-APS (SF) message, each node in the subring also performs an FDB flush operation. In particular, nodes F and G unblock their blocked ports to provide connectivity to all ring nodes. The R-APS (SF) messages disseminated from the NAFs finally arrive at interconnection nodes C and E, respectively. Then, each SF message is propagated to the other side of the subring via the R-APS virtual channel between the interconnection nodes. Meanwhile, interconnection nodes C and E additionally generate and disseminate a flush event (FE) message, called R-APS (FE), to trigger an FDB flush operation on the major ring. Since the change of the subring topology impacts traffic forwarding on the major ring, the major ring nodes A, B, and D accepting the FE message should clear their FDBs. An R-APS (SF) message is periodically transmitted from each NAF, while the failure condition persists. However, the acceptance of subsequent R-APS (SF) messages in the protection state does not retrigger an FDB flush, nor any further R-APS (FE).

2. Traffic Overshooting Problem of ERP

Whenever an active ring topology is changed by any failure or the recovery of a failure, all ring nodes should remove all learned MAC addresses in their FDBs because the previous FDB of each node is no longer valid for the new topology. Hence, the unicast data frames whose destination addresses (DAs) have not been learned are flooded over the ring until learning of the corresponding MAC addresses is completed. The duplicated frames, however, make a ring network suffer from an amount of traffic overshoot several times greater than the steady state traffic. When the volume of such flooded traffic is far greater than the link capacity, the majority of frames are lost or delayed due to queuing in a buffer. In this situation, two-fold network impairment manifests itself as extended delay and increased loss of client traffic. In addition, the burst of traffic flooding extends the address learning period. The combination of all these impairments can make protection switching and settling time greater than 50 ms. This phenomenon can be critical when a ring provides services to a large number of clients. For this reason, suitable FDB flush schemes minimizing the amount of flooded traffic should be considered. In the next section, we will introduce three schemes to guarantee less traffic overshoot than the current G.8032.

3. Flush Optimization Issue on a Multi-ring Network

The change of subring topology affects forwarding routes of traffic traversing both the major ring and subring. As shown in Fig. 1, the major ring nodes in such conditions should also perform an FDB flush to guarantee FDB consistency on a multi-ring network. In this case, the multi-ring network may suffer from far more critical traffic overshooting than a single ring because most of the nodes in the ring simultaneously broadcast data frames through the network. Any flush optimization method under a subring topology change must solve this problem. In the next section, we will also describe an efficient flush optimization technique to eliminate unnecessary FDB flush operations on the major ring.

IV. Proposed Flooding Optimization Schemes

To minimize the number of flooded frames caused by an FDB flush, we propose four protection schemes to ensure less traffic overshoot, less capacity requirement, and fast protection switching.

1. Ring-Centric FDB Flush Scheme

When an active ring topology is reconstructed by any failure of a link or node, not all FDB entries of a node actually need to be flushed. The logical topology of the subnet in which clients reside is not affected when there is a failure or recovery in a ring network, so it is not necessary to flush FDB entries associated with the client subnet. The proposed scheme, r-Flush, makes ring nodes perform an FDB flush operation on the ring ports only, excluding the client port associated with the subnet. When a failure occurs as seen in Fig. 1, all ring nodes in both rings perform an FDB flush except a port associated with its own subnet clients.

2. Port-Based FDB Flush Scheme

The second proposed scheme, p-Flush, uses the port information that receives R-APS messages related to protection switching, that is, R-APS (SF). As shown in Fig. 1, nodes G and H perform an FDB flush with respect to the failed port. As a result, the FDB entries associated with the direction opposite to the failure and the subnet of each NAF still remain. On receipt of the R-APS (SF) message, each node also performs an FDB flush operation for the entries associated with the reception port. In this scheme, the RPL owner unblocks its RPL block and flushes its FDB only if it receives the R-APS (SF) message from the non-RPL port. This restriction is necessary to prevent erroneous flush operations. Node F waits to unblock its blocked port until an SF message sent from node H arrives at its non-blocked port via the virtual channel. All nodes on the subring receive SF messages from one direction only and perform the FDB flush of the reception port solely. Once a block is removed, any subsequent SF messages do not trigger an FDB flush operation.

When interconnection nodes C and E receive the SF messages delivered from the subring, they propagate an FE message. Upon receiving the R-APS (FE) message, a major ring node performs an FDB flush for the entries associated with the port that receives the R-APS (FE) message. As seen in Fig. 1, node D conducts an FDB flush with respect to both ring ports, and nodes A and B only flush FDB entries with the reception port.

3. Extended Selective FDB Advertisement Scheme

The third proposed method, e-ADV, has been developed so that the scheme proposed in [8] can be operated on a multi-ring network. As illustrated in Fig. 2, the e-ADV makes nodes performs the r-Flush by which FDB entries associated with the ports connected to a ring are removed.

In addition, the e-ADV lets all ring nodes exchange their remaining FDB entries with each other. This helps ring nodes learn the subnet clients of each node in a short period of time. Each ring node generates a subnet address list (SAL) containing DAs of the FDB entries associated with its own subnet. The generated SAL is conveyed in the payload of the

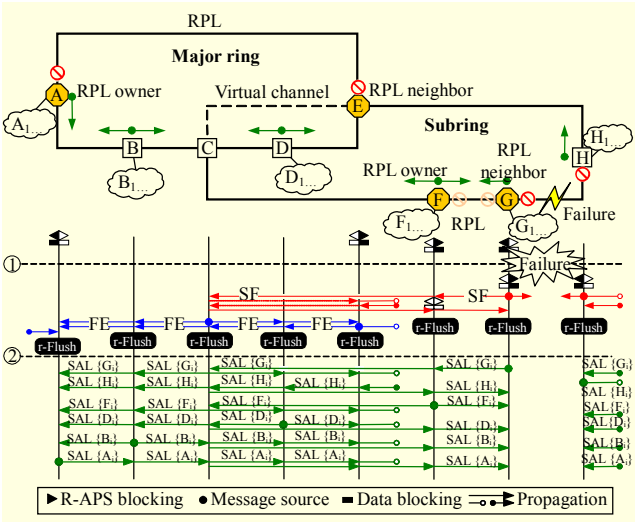

Fig. 2. Protection switching in extended selective FDB advertisement.

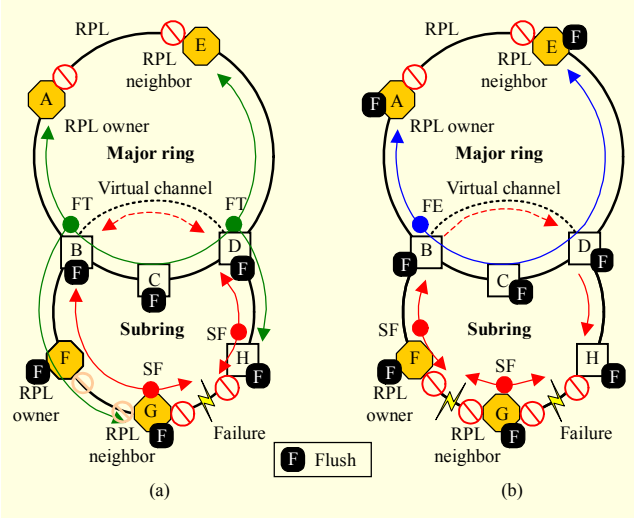

Fig. 3. Flush optimization by flush triggering under (a) the first failure condition and (b) the second failure condition.

R-APS (SAL), which is newly defined to support the proposed scheme. If the size of the SAL is too large to fit the maximum transmission unit size, multiple R-APS (SAL) messages with fragmented-payload will result. Then, the R-APS (SAL) message is transmitted through either both ring ports or a single ring port depending on whether it is an NAF. To prevent FDB inconsistency in case of a unidirectional failure, an NAF should send the R-APS (SAL) message in the direction of an unblocked ring. Thus, nodes G and H multicast their own R-APS (SAL) messages containing $\{G_i\}$ and $\{F_i\}$, respectively, only to the non-failed ring port. Note that $\{X_i\}$ is a set of client addresses attached to a ring node X. Meanwhile, other ring nodes except the NAFs multicast their R-APS (SAL) messages over both ring ports. The RPL owner and RPL neighbor in a normal state do not send or relay R-APS (SAL) messages to the RPL port.

Upon receipt of an R-APS (SAL) message, each node copies the message and sends the original to the next node as defined in the node model of the G8032 recommendation. The node then learns the MAC addresses indicated in the SAL of the copied frame with the R-APS (SAL) reception port. This operation, the indirect MAC address learning process, enables a ring node to build the FDB as if it received multiple individual data frames. When all ring nodes conduct this process of all received SALs, protection switching is completed.

4. Flush Triggering Scheme for Interconnect Rings

When there is a topology change on a subring, the FDB flush operation does not have to occur in all nodes on the major ring. Instead, only the nodes on the active path between two interconnection nodes need to be flushed. In Fig. 1, nodes A and B on the major ring actually need not perform an FDB flush operation. As the RPL blocks data traffic under normal conditions, communications between either node A or node B and a node on the subring are achieved along one direction from the view of node A or B. Using this fact, the last proposed scheme called flush triggering defines an extra message called flush trigger (FT).

Flush triggering mechanism under subring topology change is illustrated in Fig. 3. In Fig. 3(a), interconnection nodes B and D receiving the SF message transmitted from subring nodes perform an FDB flush and generate R-APS (FT) messages. Upon reception of R-APS (FT), each node does not perform an FDB flush immediately. A node performs an FDB flush operation only if the FT messages arrive from both directions within a given time. For this reason, only node C clears its FDB, but nodes A and E do not.

When an interconnection node generates the FT message by subring topology change, it should check in which state its own subring is. As shown in Fig. 3(b), if the subring is in a protection state, the interconnection node must generate a regular FE message, R-APS (FE). As the SF message from the second failure cannot arrive at both interconnection nodes, the FT message cannot cause the major ring nodes to perform a flush.

This scheme is used with the three schemes proposed in our performance evaluation study.

V. Evaluations

1. Simulation Scenario

The protection switching performance by the FDB flush, the r-Flush, the p-Flush, and the e-ADV was evaluated using the OPNET simulator [15]. To observe how transient traffic under

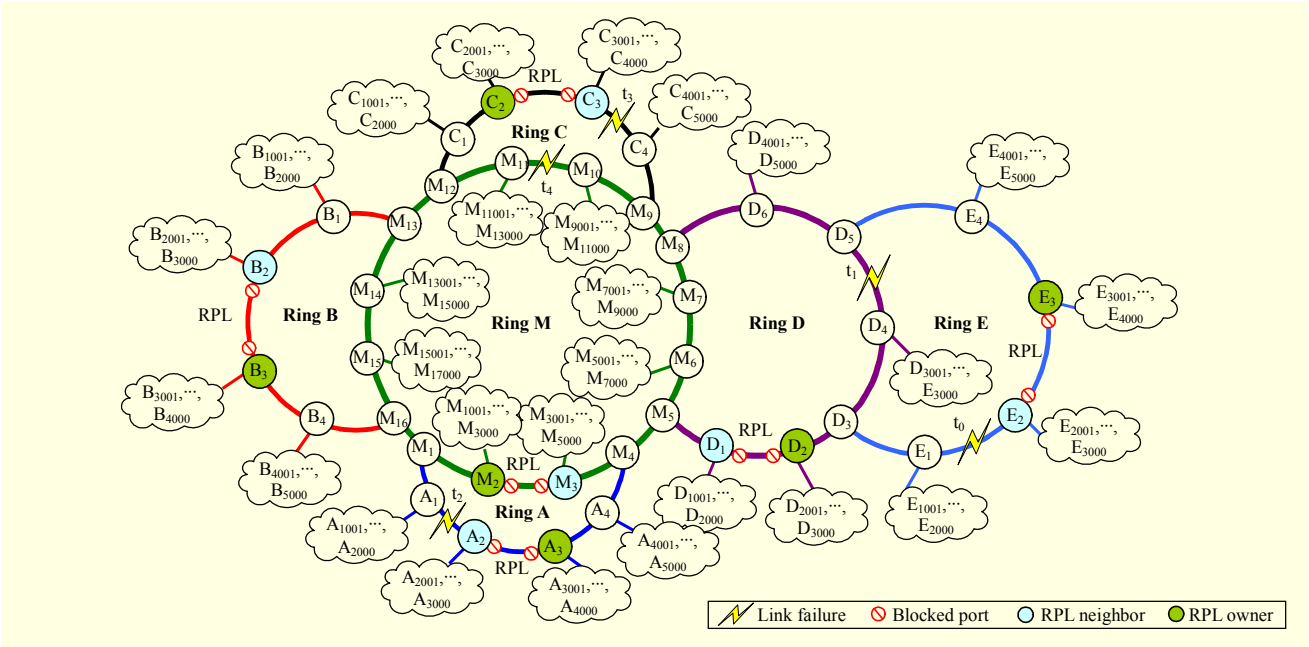

Fig. 4. Simulation scenario.

failure conditions affects the performance of a multi-ring network, the link utilizations of each ring are measured every 4 ms. As shown in Fig. 4, the simulation scenario is modeled as a multi-ring network where six interconnected rings are spread in a metro-scale area. Ring M is on top of the multi-ring and 1,200 km in circumference. It consists of sixteen Ethernet ring nodes (M_1 to M_{16}). In this ring, each node is connected to two adjacent ring nodes with an 80 km 10 Gbps full-duplex link (200 μs time-of-flight). Every node except interconnection nodes is assumed to have one client subnet using a 1 Gbps link where 2,000 clients reside. Subring A (nodes A_1 to A_4), subring B (nodes B_1 to B_4), subring C (nodes C_1 to C_4), and subring D (nodes D_1 to D_6) are interconnected to major ring M with multiple shared links through the use of two interconnection nodes grouped in ring M. Subring E with four nodes (E_1 to E_4) then forms another subring layer and is attached to subring D. Each node of every subring is also connected to two neighboring nodes with a 40 km 1 Gbps links and single subnet with 1,000 clients. Each client exponentially generates 80 kbps traffic with 4,000 bits/frame toward destinations equally distributed among all other clients in the multi-ring network. Meanwhile, every subring owns a virtual channel between two interconnection nodes for the subring to support R-APS connectivity. The RPL of each ring is initially blocked to ensure a loop-free ring topology.

To evaluate how well the proposed schemes deal with protection switching on a multi-ring network, we carried out our experiments focusing on the following three aspects: performance by subring topology change, performance comparison due to different locations of the active paths on upper rings, and performance of the major ring. We describe the results for each of these experiments in the following subsections. Of those proposed, the e-ADV in particular assumes that an R-APS (SAL) message contains at most 200 MAC addresses. As the list-based R-APS frame conveys lengthy information compared to other normal R-APS frames, its processing time is set to be the sum of the service time of one normal R-APS frame and the processing time of all the address entries contained in a list, where the processing time of one entry is set to half of the service time per data frame as a data frame normally experiences two FDB accesses, namely, DA lookup and source address (SA) learning.

2. Protection Switching under Subring Topology Change

When link E_1-E_2 of subring E fails at 2.0 s (= t_0), upper rings D and M perform an FDB flush. To observe how the subring topology change affects upper rings, we monitored the utilization of links where the occurrence of flooded frames is most frequent in each ring. These results are shown in Figs. 5 and 6. The figures also include protection switching behavior with flush triggering to demonstrate its effectiveness. We also examined how different rates of data and R-APS frames affect the protection performance of the proposed schemes with or without flush triggering. The upper and lower bounds of the service rates of two types of frames are shown in Table. 1. For the upper bound, both data and R-APS service rates are set to be double the lower bound. The lower bound of the data frame

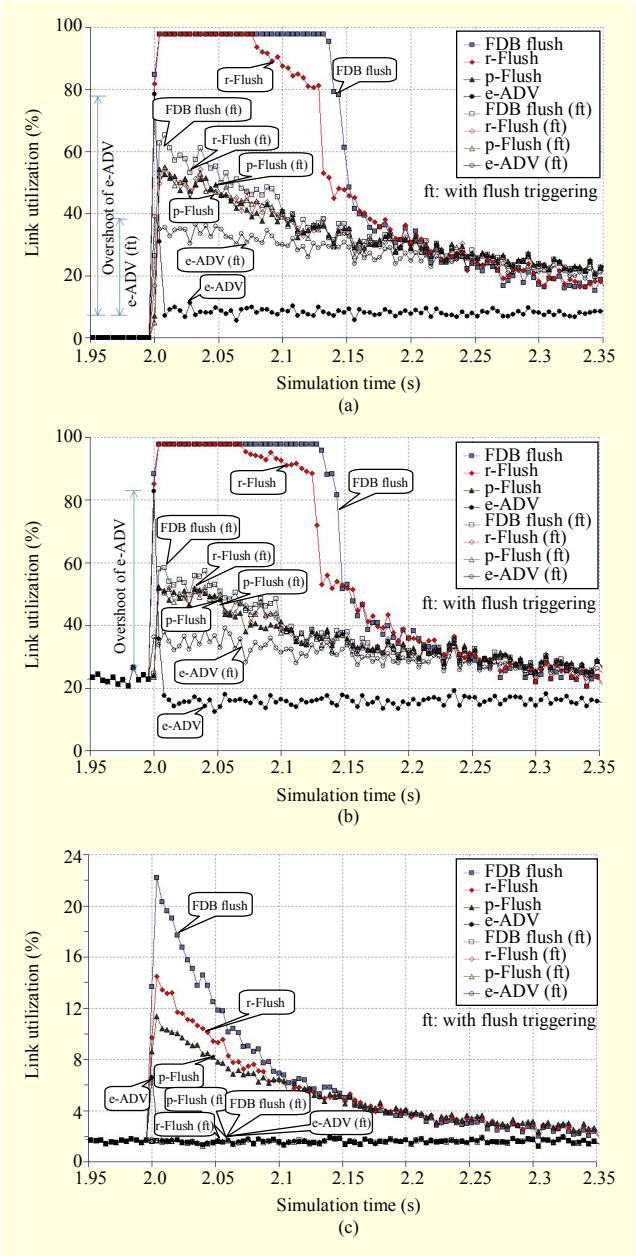

Fig. 5. Utilization (a) at 1 Gbps link $E_3 \rightarrow E_2$, (b) at 1 Gbps link $D_4 \rightarrow D_3$, and (c) at 10 Gbps link $M_4 \rightarrow M_3$ in upper bound scenario.

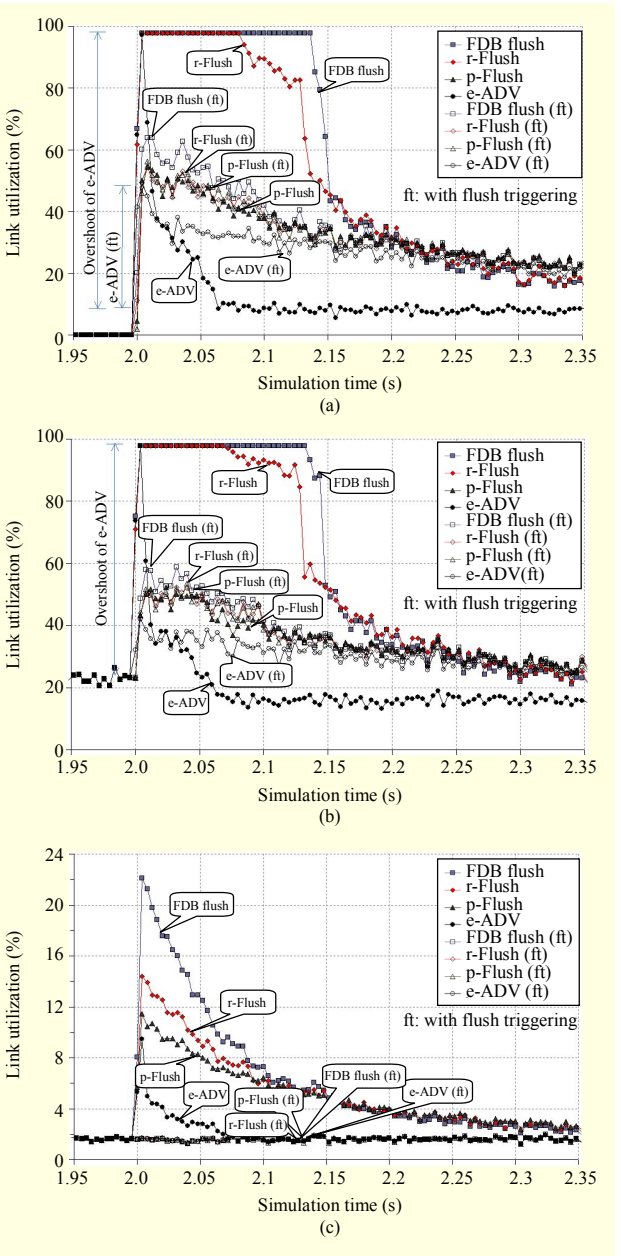

Fig. 6. Utilization (a) at 1 Gbps link $E_3 \rightarrow E_2$, (b) at 1 Gbps link $D_4 \rightarrow D_3$, and (c) at 10 Gbps link $M_4 \rightarrow M_3$ in lower bound scenario.

service rate is chosen to be the lowest value at which data frames do not suffer from buffering in the normal state. The R-APS frame service rate is set to be larger than the minimum value to process all the R-APS (SAL) frames within 50 ms. Note that the number of the R-APS (SAL) frames generated over the entire network in a failure event was about 200.

As shown in Fig. 5(a), after the failure at t_0, the FDB flush without flush triggering experiences full link utilization for almost 150 ms. Transient traffic is not stabilized even after 300 ms. When the flush triggering technique is applied, the

Table 1. Service rate of data and R-APS frames.

Service rate per ring network		Lower bound (mpps)	Upper bound (mpps)
Data frame	10 G	7.5	15
	1 G	0.75	1.5
R-APS frame	10 G	0.05	0.1
	1 G	0.005	0.01

transient traffic of subring E does not experience the saturated utilization. However, it still shows the heaviest transient traffic among the proposed methods with flush triggering. The r-Flush without flush triggering provides less traffic overshooting than the original FDB flush, but it also experiences the full utilization for about 80 ms. When flush triggering is added, it significantly reduces the amount of flooded traffic as in FDB flush. Next, p-Flush further reduces the number of flooded frames compared to the previous two schemes. There is no significant difference between the results with or without flush triggering. However, p-Flush shows slightly more link utilization than FDB flush or r-Flush after 200 ms. It was observed from a separate experiment, which is not shown in this paper, that the number of learned SA entries for two previous schemes is larger than the number of learned SA entries for p-Flush because of more duplicated unknown DA frames up to 200 ms. Consequently, convergence of p-Flush takes slightly more time than FDB flush or r-Flush. Finally, e-ADV provides lower traffic overshoot and faster stabilization than the other schemes. Without flush triggering, e-ADV generates just one spike of traffic for only a few milliseconds and immediately reaches a steady state. This spike jumped up to 80% of utilization but was immediately suppressed due to the indirect MAC learning from the propagated SAL frames. However, e-ADV with flush triggering reveals some transient traffic. This is because the upper ring nodes that are not on the active path do not perform an FDB flush operation or generate their own subnet address list. The nodes on the active path flush their FDBs and broadcast data frames with unknown destinations until they learn the location of corresponding clients. In spite of such a partial exchange of subnet address lists, e-ADV guarantees the most rapid and stable protection switching among all the schemes with flush triggering.

The link utilization in subring D, which is an upper ring of subring E, is shown in Fig. 5(b) and shows the similar shape to Fig. 5(a). The e-ADV scheme also guarantees more rapid and stable protection switching than the other schemes.

The performance of the proposed methods is clearly distinguished at the link $M_4 \rightarrow M_3$ because a 10 Gbps link is enough to accommodate the traffic overshoot. As seen in Fig. 5(c), the effectiveness of flush triggering is not apparent because the link $M_4 \rightarrow M_3$ is not affected by the failure at link E_1-E_2. The R-APS (FT) message on the major ring is generated only from M_8 due to the location of the block on subring D. Consequently, none of the nodes on the major ring flush their FDBs. In Fig. 5(c), the FDB flush presents the largest traffic overshoot. At the peak, the link utilization of link $M_4 \rightarrow M_3$ is about 22%. This value is nearly fifteen times the steady state value. Also, this transient traffic lasts for about 300 ms. In the case of r-Flush, the traffic overshooting increases up to ten times the steady state. The p-Flush scheme further reduces this occurrence over r-Flush, but it also converges slowly. The peak utilization was nearly seven times greater than the steady state. As expected, e-ADV still guarantees the lowest traffic overshooting and the most stable protection switching among all flush schemes.

We analyzed the results of link utilization monitored in the case of the lower bound. As a whole, the simulation results of the proposed methods show a strong resemblance to appearances in the upper bound except e-ADV without flush triggering. Due to the longer processing time, e-ADV without flush triggering experiences transient traffic even after 50 ms in all cases of Fig. 6. The spike of traffic reached full utilization in the subrings D and E. Nonetheless, e-ADV shows the fastest protection switching performance. Actually, the service rates indicated in the lower bound are a little far away from those of a commercial metro Ethernet switch product. This demonstrates that e-ADV guarantees an effective protection solution with the minimal capacity requirement for practical Ethernet ring networks.

In addition to the failure at t_0, another subring link failure occurs at link D_4-D_5 of subring D at 4.0 s (=t_1). As the observed results were similar to those of the failure at t_0, we omit its performance results in this paper.

3. Comparison of Protection Switching Performance Due to Different Locations of Active Paths on Upper Rings

As mentioned in section IV.4, the optimality condition under subring topology change makes a multi-ring network avoid unnecessary FDB flush operations by using active path information. To explore the performance difference, we examined the case in which a failure occurs at each of two subrings, whose active paths on the major ring are different. Failures are assumed to occur at 6.0 s (= t_2) and 8.0 s (= t_3), and the upper bound scenario is assumed. First, we monitored link $M_4 \rightarrow M_3$ in which the biggest traffic overshooting occurs in the case of the failure of subring A. Second, we observed link $M_{10} \rightarrow M_{11}$ which is located at a symmetrical position against the RPL link of ring M. This link is the location which suffers from the heaviest transient traffic under a failure in both subrings A and C.

When link A_1-A_2 of subring A fails at 6.0 s, the R-APS (SF) messages disseminated from nodes A_1 and A_2 individually arrive at interconnection nodes M_1 and M_4. Then, nodes M_1 and M_4 propagate the R-APS (FT) message over both ring ports. Since the blocked link is placed within the shared links between the interconnection nodes, the active path of subring A consists of all of the links except the shared links. Hence, fourteen nodes on the active path perform an FDB flush. As

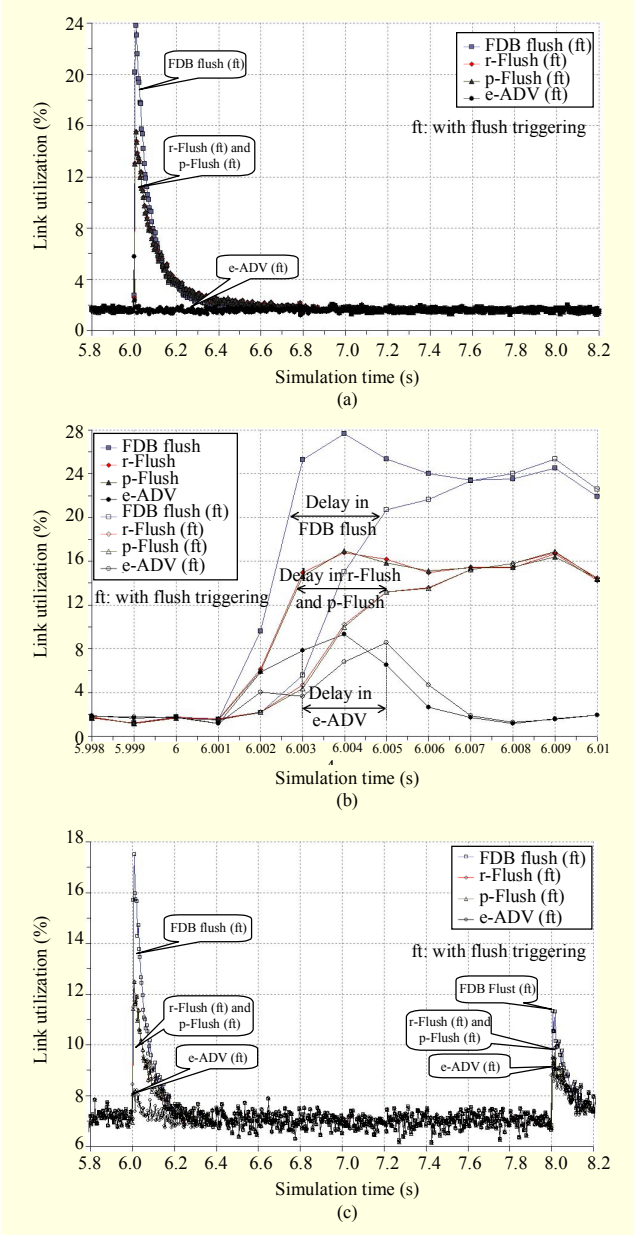

Fig. 7. (a) Utilization at link $M_4 \rightarrow M_3$, (b) switching delay difference at link $M_4 \rightarrow M_3$, and (c) utilization at link $M_{10} \rightarrow M_{11}$ with flush triggering.

shown in Fig. 7(a), the FDB flush scheme overshoots sixteen times greater than the steady state. Next, r-Flush and the p-Flush ensue. The r-Flush and the p-Flush schemes show the exact same performance because both methods flush the same entries from the FDB of the node on active path. The traffic overshoot increased up to ten times more than that in the normal state. Finally, e-ADV shows the fastest protection switching while producing the small amount of transient traffic.

Flush triggering provides an optimized FDB flush under a subring topology change. However, it may cause delayed protection switching because flushing FDB is triggered by two R-APS (FT) messages received at both ring ports. To investigate the protection switching delay, we monitored link $M_4 \rightarrow M_3$ at every one millisecond interval wherein the delay of flush triggering by the R-APS (FT) message can be revealed. As shown in Fig. 7(b), every scheme with flush triggering delays the FDB flush for about 2 ms to 3 ms. Compared to the 50 ms protection switching requirement, the duration of the delay by flush triggering is relatively small. The different link utilization values between Fig. 7(a) and Fig. 7(b) are attributed to the different measurement interval.

When link C_2-C_3 of subring C fails at 8.0 s, the active path is equal to the shared links between subring C and ring M. The length of the active path is relatively short compared to that in subring A, and only four nodes including interconnection nodes M_9 and M_{12} perform an FDB flush. Therefore, the amount of overshoot traffic is far less than the previous case of the failure on subring A. As shown in Fig. 7(c), the FDB flush shows the largest transient traffic, but it is similar to that of r-Flush and p-Flush. The e-ADV scheme reveals transient traffic for about 200 ms because twelve major ring nodes except the flush nodes do not propagate their SAL information. Meanwhile, since the nodes on inactive paths do not trigger an FDB flush, most of the flooded traffic is filtered at nodes M_8 and M_{13}. As a result, link $M_4 \rightarrow M_3$ in Fig. 7(a) displays no traffic overshooting after 8.0 s.

4. Performance of Protection Switching in Major Ring

Finally, we evaluated how each of the proposed schemes handles transient traffic caused by protection switching under a link failure on the major ring. Since the flush triggering method is only applicable to the case of the subring topology change, it is not considered. First, we observed the utilization of the link $M_8 \rightarrow M_9$ after the failure of the link M_{10}-M_{11} at 10 s (= t_4) because the link M_{10}-M_{11} carries the heaviest traffic among all links before failure. In Fig. 8, the FDB flush scheme introduces the largest traffic overshoot because it removes all FDB entries at the failure event. At the peak, the link utilization of link $M_8 \rightarrow M_9$ surges up to three times more than that in the normal state. The overshoot lasts for about 100 ms, and transient traffic cannot reach the steady state even after 300 ms. Meanwhile, r-Flush significantly reduces the overshoot. By doing the FDB flush excluding the subnet, it only requires a link capacity less than half that of the steady state before the failure. Next, p-Flush further eliminates the flooded frames through an FDB flush operation using the information of the port where the R-APS (SF) message is received. At the peak, it reaches up to about 130% of the steady state value. Last, the e-ADV shows that the amount of traffic overshoot is far less than with other

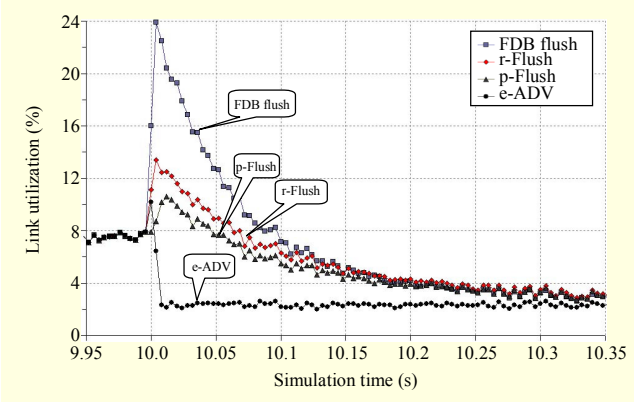

Fig. 8. Utilization at 10 Gbps link $M_8 \rightarrow M_9$.

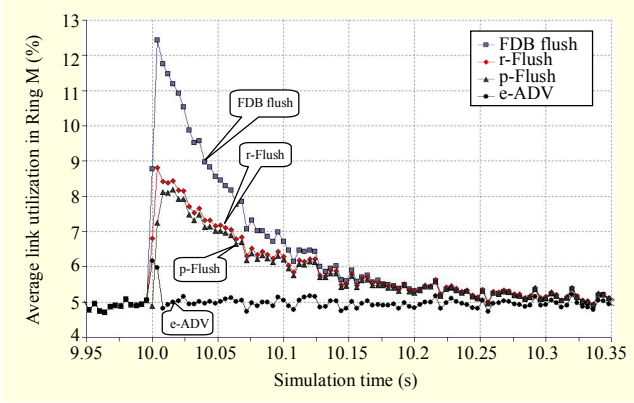

Fig. 9. Average utilization on 10 Gbps Ring M.

schemes. As each node exchanges its subnet FDB entries, it immediately removes the traffic overshoot and stabilizes transient traffic within about 10 ms.

Second, we obtained the average link utilization on the major ring to evaluate how much transient traffic each scheme produces overall. As shown in Fig. 9, FDB flush reveals the most flooded frames and reaches about 250% of the normal state at the peak. The overshoot phenomenon lasts for about 300 ms. Meanwhile, both r-Flush and p-Flush schemes significantly reduce the transient traffic. The two schemes are indistinguishable in overall performance. The e-ADV shows the lowest link utilization of all the schemes. The average utilization increases up to 120% of the normal state, and the transient traffic lasts about 10 ms.

5. Summary and Arguments

The ring-centric FDB flush makes ring nodes prevent flushing FDB entries attached to subnets. Without any complicated operation, it simply provides far less transient traffic than the original FDB flush.

The port-based FDB flush guarantees less traffic overshoot than the ring-centric FDB flush as it reduces unnecessary flushed FDB entries. However, it has a drawback that the RPL owner keeps the block until the R-APS (SF) messages arrive at non-RPL port, and this can cause a protection switching delay. Moreover, this scheme cannot be used in a multi-ring network where there is no R-APS virtual channel. If a subring without a virtual channel fails, several FDB entries will never be flushed due to fragmentary R-APS connectivity. Furthermore, the flush logic defined in G.8032 uses an FE timer to prevent duplicated flush triggering. For port-based FDB flush operations, this function also has to be disabled. In the next version of G.8032 recommendation, when a node receives an R-APS message with a pair of node ID and blocked port number, an FDB action is triggered if the pair information is different from that of the previous pair stored at the reception port and the pair already stored at the other port. In other words, a node usually performs an FDB flush twice under a protection condition. This means that the port-based FDB flush cannot be differentiated from the ring-centric FDB flush. In conclusion, the port-based FDB flush seems difficult to justify in the version of G.8032 currently being considered.

The third scheme, the extended selective FDB advertisement, provides the most realistic and rapid failover because the indirect MAC address learning process makes ring nodes build the FDB entries right after failure. It shows the most reliable performance among all FDB flush schemes, even under unexpected circumstances, such as processing delay or loss of the R-APS (SAL), it does not introduce any malfunctioning in the ring networks. Moreover, the newly defined R-APS (SAL) message is relayed with no modification in an intermediate node, resulting in faster protection switching. In addition, the FDB advertisement is designed to be compatible and interoperable with the standard FDB flush scheme without any additional operation. In the environment of the mixture of ring nodes with the extended selective FDB advertisement and the FDB flush, correct protection switching and ring operation are guaranteed. Since it fully complies with the node architecture and R-APS protocol message relay model defined in G.8032 recommendation, it can be an effective protection solution for Ethernet ring networks.

Finally, we investigated the effects of the combination of the flush triggering scheme and each of the three proposed schemes. When flush triggering is combined with the ring-centric FDB flush or the port-based FDB flush, it significantly helps reduce transient traffic. However, a combination with extended selective FDB advertisement scheme causes rather unstable protection switching results. In conclusion, the extended selective FDB advertisement guarantees the fastest and the most stable protection switching even without the flush optimization scheme.

Table 2. Stabilization time of the proposed schemes.

Features		FDB flush	r-Flush	p-Flush	e-ADV
T_{stable}	Up	≤300 ms	≤300 ms	≤300 ms	≤10 ms
	Low	≤ 300 ms	≤300 ms	≤300 ms	≤60 ms

In Table 2, the stabilization time of each of the four approaches is summarized with respect to the upper and lower bound scenarios.

VI. Conclusion

In this paper, network stability issues in Ethernet ring protection switching are discussed. To minimize the transient traffic overshoot caused by flooded traffic, we proposed three flush optimization schemes and one technique using the optimality condition in which the nodes on non-active paths do not need to perform an FDB flush operation. Experiment results showed how well the proposed schemes handle transient traffic on a multi-ring network. Among the proposed methods, the extended FDB advertisement provided the fastest and most stable protection switching. To achieve sub-50 ms protection switching time, it is crucial to utilize an appropriate FDB flush optimization scheme for G.8032 ERP networks.

References

[1] K. Fouli and M. Maier, "The Road to Carrier-Grade Ethernet," *IEEE Commun. Mag.*, Mar. 2009, vol. 47, no. 3, pp. S30-S38.

[2] IEEE Std. 802.1D, "IEEE Standard for Local and Metropolitan Area Networks – Media Access Control (MAC) Bridge," June 2004.

[3] IEEE Std. 802.1w, "IEEE Standard for Local and Metropolitan Area Networks, Amendment 2: Rapid Reconfiguration for Spanning Trees," June 2001.

[4] IEEE Std. 802.1s, "IEEE Standard for Local and Metropolitan Area Networks - Virtual Bridged Local Area Networks, Amendment 3: Multiple Spanning Trees," Dec. 2002.

[5] IEEE Std. 802.17, "Part17: Resilient Packet Ring (RPR) Access Method and Physical Specifications," 2004.

[6] IEEE Std. 802.17b, "Part17: Resilient Packet Ring (RPR) Access Method and Physical Specifications – Amendment 1: Spatially Aware Sublayer," 2007.

[7] ITU-T Rec. G.8032/Y.1344, "Ethernet Ring Protection Switching," 2008.

[8] K. Lee, J. Ryoo, and S. Min, "An Ethernet Ring Protection Method to Minimize Transient Traffic by Selective FDB Advertisement," *ETRI J.*, vol. 31, no. 5, Oct. 2009, pp. 631-633.

[9] J. Ryoo et al., "Ethernet Ring Protection for Carrier Ethernet Networks," *IEEE Commun. Mag.*, vol. 46, no. 9, Sept. 2008, pp. 136-143.

[10] J.K. Rhee, J. Im, and J. Ryoo, "Ethernet Ring Protection using Filtering Database Flip Scheme for Minimum Capacity Requirement," *ETRI J.*, vol. 30, no. 6, Dec. 2008, pp. 874-876.

[11] A. Kvalbein, S. Gjessing, and F. Davik, "Performance Evaluation of an Enhanced Bridging Algorithm in RPR Networks," *The 3rd IEEE Int. Conf. Networking (ICN)*, 2004, pp. 760-767

[12] P. Setthawong and S. Tanterdtid, "Flood Avoidance Mechanisms for Bridged Resilient Packet Rings," *J. Computer Science and Technol.*, vol. 23, no. 5, Sept. 2008, pp. 815-824.

[13] W. Zhong et al., "Optical Resilient Ethernet Rings for High-speed MAN Networks [invited]," *J. Optical Networking*, vol. 4, no. 12, Dec. 2005, pp. 784-806.

[14] J. Ryoo et al., "OAM and Its Performance Monitoring Mechanisms for Carrier Ethernet Transport Networks," *IEEE Commun. Mag.*, vol. 46, no. 3, Mar. 2008, pp. 97-103.

[15] OPNET Technologies Inc. http://www.opnet.com.

Kwang-Koog Lee received his BS and MS degrees in electronic communication engineering from Kangwon National University, Chuncheon, Rep. of Korea, in 2006 and 2008, respectively. He is currently working towards his PhD degree in broadband network technology at the University of Science and Technology, Rep. of Korea. Since 2008, he also joined ETRI, Rep. of Korea, where he has worked on switching and management technologies for high-speed optical transmission systems. His research interests are in parallel and distributed computing, and optimized protocol design and performance analysis in wired/wireless networks.

Jeong-dong Ryoo is a principal member of research staff in ETRI, Rep. of Korea. He holds MA and PhD degrees in EE from Polytechnic University, Brooklyn, NY, USA, and a BEE from Kyungpook National University, Rep. of Korea. After completing his PhD study in telecommunication networks and optimization, he started working for Bell Labs, Lucent Technologies, NJ, in 1999. While he was with Bell Labs, he was mainly involved with performance analysis, evaluation, and enhancement study for various wireless and wired network systems. Since he left Bell Labs and joined ETRI in 2004, his work has been focused on next generation network and carrier class Ethernet technology research, especially participating in OAM and protection standardization activities in ITU-T. He co-authored *TCP/IP Essentials: A Lab-Based Approach* (Cambridge University Press, 2004). He is a member of Eta Kappa Nu association.