Faultless Protection Methods in Self-Healing Ethernet Ring Networks

Kwang-koog Lee, Jeong-dong Ryoo, and Bheom Soon Joo

Self-healing Ethernet rings show promise for realizing the SONET/SDH-grade resilience in Carrier Ethernet infrastructures. However, when a ring is faulty, high-priority protection messages are processed in less time than low-priority data frames are processed. In this situation, any delayed data frames either being queued or traveling through long ring spans will cause the ring nodes to generate incorrect forwarding information. As a result, the data frames spread in the wrong direction, causing the ring to become unstable. To solve this problem, we propose four schemes, that is, dual flush, flush delay timer setting, purge triggering, and priority setting, and evaluate their protection performance under various traffic conditions on a ring based on the Ethernet ring protection (ERP) method. In addition, we develop an absorbing Markov chain model of the ERP protocol to observe how traffic congestion can impact the protection performance of the proposed priority setting scheme. Based on our observations, we propose a more reliable priority setting scheme, which guarantees faultless protection, even in a congested ring.

Keywords: Carrier Ethernet, Ethernet ring protection, absorbing Markov chain, filtering database.

I. Introduction

With the rapid growth of high bandwidth service in recent years, Ethernet has evolved into a highly dominant carrier-grade network technology [1]. As Ethernet has progressed considerably among the transport networks, it has been challenged by service providers who need rapid and reliable recovery capabilities to guarantee the availability of their services.

Originally focusing on the scope of LAN, Ethernet relied on the spanning-tree protocol (STP) to perform forwarding to ensure loop avoidance [2]. However, the STP approach neither acts fast upon any topology change nor ensures optimal forwarding paths. To mitigate these drawbacks, two enhanced STP-based approaches, that is, rapid spanning tree protocol and multiple spanning tree protocol, were proposed [3], [4]. Yet, the convergence time for each is still too long to meet the sub-50-ms protection switching time requirement. To replace the STP-based approaches, several self-healing Ethernet ring networks, such as resilient packet ring (RPR) [5], Ethernet ring protection (ERP) [6]-[8], and Ethernet automatic protection switching (EAPS) [9], emerged as standardization alternatives, and several vendor-specific self-healing rings were also introduced [10], [11]. These ring-based approaches do not require complicated control, provisioning overheads, or excessive information exchange to achieve the SONET/SDH-grade resilience under a link or node failure.

In general, all the aforementioned self-healing Ethernet ring networks except RPR work on the basis of a filtering database (FDB) flush operation, which forces ring nodes to remove all learned media access control (MAC) addresses after a topology change. This operation is essential for the ring networks because the ring nodes can maintain consistent forwarding

Manuscript received May 10, 2012; revised July 7, 2012, accepted July 19, 2012.

Kwang-koog Lee (phone: +82 10 2958 8179, kwangkooglee@gmail.com) is with the Department of Computer Science, Korea Advanced Institute of Science and Technology, Daejeon, Rep. of Korea.

Jeong-dong Ryoo (corresponding author, ryoo@etri.re.kr) and Bheom Soon Joo (bsjoo@etri.re.kr) are with the Advanced Communications Research Laboratory, ETRI, Daejeon, Rep. of Korea.

http://dx.doi.org/10.4218/etrij.12.1812.0102

information even when a ring topology is changed by a failure, recovery of the failure, or such operator commands as "manual switch" and "forced switch." However, the FDB flush action is triggered by a high-priority protection message that is forwarded by a ring node in less time than low-priority data frames are forwarded. In this situation, any delayed data frames either being queued or traveling through long ring spans will arrive at some ring nodes after their flush operations are completed, causing the ring nodes to build erroneous forwarding information. As a result, subsequent data frames referring to such FDB entries at the ring nodes are forwarded in the wrong direction, and the ring network ultimately becomes unstable. In this regard, stabler protection mechanisms to guarantee the consistency of the FDB should be considered.

In this paper, we propose four schemes to remove the faulty FDB entries generated from the delayed data frames: dual flush, flush delay timer setting (f-timer), purge triggering (purging), and priority setting (set prio). First, the dual flush scheme lets every ring node refresh its FDB table with respect to each of its two ring ports. In other words, the FDB flush operation is processed at least twice at a ring node. Next, the f-timer makes ring nodes prolong the FDB flush operation until the end of its timer period. The purging scheme is used to purge all the buffered data frames along with the FDB flush. Finally, the set prio scheme sets the priority of protection messages to be the same as the lowest priority of the data frames. To observe how well each scheme achieves the faultless protection in a failure event, we evaluate their protection performance using OPNET [21] under various traffic conditions on an Ethernet ring based on the G.8032 ERP method. The simulation results show that our solutions effectively remove the erroneous FDB entries.

In addition, we observe that the set prio scheme may not guarantee successful protection in severe congestion conditions because the protection messages can be dropped due to the lowered priority. To investigate how seriously the protection messages with the low priority are affected by congestion, we develop an absorbing Markov chain model of the ERP protocol and theoretically estimate the lower bound of the expected protection switching time. Assuming the ring nodes are equipped with the M/M/1/k queue, our analytic results find that a sufficient burst of protection messages is necessary to overcome the loss of the protection messages. Based on this observation, we suggest an enhanced priority scheme to find the optimal number of protection messages. We also conduct simulations to validate the results of our developed model. The simulation results demonstrate that our analytic model is indeed capable of estimating the expected protection switching time.

The remainder of this paper is organized into seven sections. Section II introduces the basic principles of ring-based Ethernet protection technologies and describes each scheme in detail. In section III, we define the FDB inconsistency problem that occurs in self-healing rings and introduce related works. Then, we provide detailed descriptions of our proposed schemes in section IV, and their performances are evaluated and discussed in section V. Section VI develops an analytic model using the absorbing Markov chain. The enhanced set prio is introduced in section VII. Finally, the conclusion is drawn in section VIII.

II. Ring-Based Ethernet Protection Technologies

1. Flush-Based Ring Protection Mechanism

Ethernet-based self-healing ring networks, such as RPR, ERP, EAPS, rapid ring protection protocol (RRPP) [10], and resilient Ethernet protocol (REP) [11], have been developed as substitutes for the STP-based approaches. Except for RPR, the self-healing Ethernet ring technologies have similar protection behaviors under a failure condition. Their detailed procedures are depicted in Fig. 1 using an example of a six-node Ethernet ring.

In Fig. 1(a), each ring node is connected to two adjacent nodes, and node A is designated to the master node, which is responsible for blocking its one ring port to create a logical loop-free topology in the normal state of the ring. When link B-C bidirectionally fails, as depicted in Fig. 1(b), the nodes adjacent of the link failure (NAFs), nodes B and C, detect the failure condition and block their port where the failure is detected. Then, each of them generates and transmits a protection message, notifying the other ring nodes of the failure situation. This message is multicast along both directions. Upon receiving this message, each ring node recognizes the failure condition of its ring and flushes its FDB table to eliminate invalid MAC addresses. In particular, the master node additionally removes its logical block to recover the connectivity to all ring nodes. Finally, the ring transitions to a stable protection state.

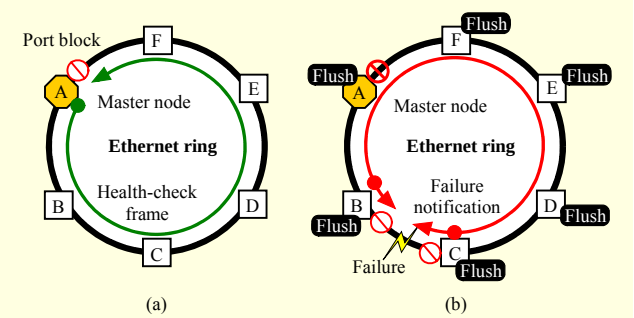

Fig. 1. Basic operations of self-healing Ethernet ring: (a) in normal state and (b) in protection state.

2. Ring-Based Ethernet Protection Technologies

ERP, as defined by ITU-T G.8032, defines one special link called "the ring protection link (RPL)." A ring node adjacent to the RPL, which is called "the RPL owner," blocks a ring port attached to the RPL to provide a loop-free structure in a normal ring. In conjunction with the continuity check functionality defined in IEEE 802.1ag [12] or ITU-T Y.1731 [13], ERP can detect a fault condition on an Ethernet ring link. When a link failure occurs, the NAFs immediately block the port in the direction that the failure has occurred and transmit the ring automatic protection switching (R-APS) (signal fail [SF]) messages in both directions. Upon receiving an R-APS(SF) message, a ring node clears its FDB table. In particular, the RPL owner unblocks the RPL port so as to provide connectivity to all ring nodes. ERP defines an R-APS protocol to coordinate these protection activities. It is added on Ethernet OAM protocol data unit [14].

EAPS and RRPP designate a single node in an Ethernet ring as the master node. All other nodes on the ring are referred to as transit nodes. One port of the master node is assigned as the primary port and the other is the secondary port. The master node blocks the secondary port in normal operation. Once a transit node detects the link down on any of its ring ports, it sends a link-down control message to the master node in its EAPS (or RRPP) domain. Then, the master node unblocks its secondary port and transmits the ring-down-flush-FDB (or complete-flush-FDB in RRPP) message notifying the transit nodes of the changed topology over both ring ports. A node accepting this message refreshes its FDB table.

Cisco designed the Resilient Ethernet Protocol (REP) technology to meet the requirements for fast and predictable convergence in layer 2 ring topologies. The REP also has at least one logical block in a given domain (segment), which is called "the alternate port." When a failure occurs, nodes with the failed port propagate the failure notification message to all REP peers. Subsequently, the REP opens the alternate port and allows data traffic to be forwarded. Every peer also performs an FDB flush.

III. Stability Problem of Flush-Based Self-Healing Rings

1. Filtering Database Inconsistency Problem

To accomplish fast protection switching in the case of a failure, ring nodes with any of the protection technologies generally forward their own protection messages in less time than it takes to process the data frames. Thus, the protection messages usually have a higher priority than any other frames, such as the data frames. However, this configuration can cause

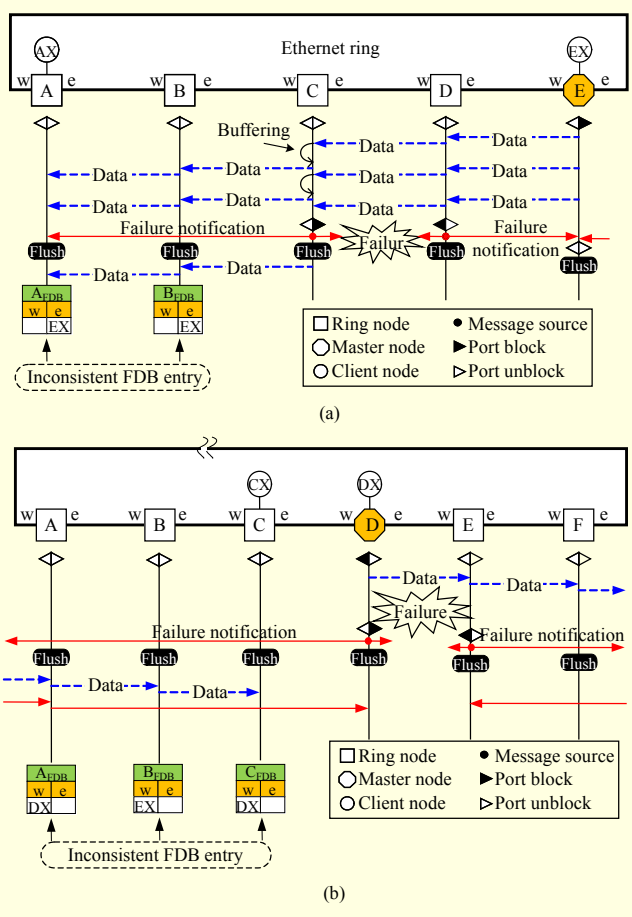

Fig. 2. FDB inconsistency problem (a) in congested ring and (b) in noncongested ring.

the ring nodes to build erroneous FDB information. Such an undesirable phenomenon is classified in one of two ways: a congested ring or a noncongested ring. Figures 2(a) and 2(b) illustrate a congested ring and a noncongested ring, respectively.

Case 1. Suppose data frames are generated from client EX and are forwarded to client AX. We also assume link C→B is currently congested. Therefore, node C continues to buffer data frames forwarded from node C, D, or E in its west ring port (w). When link C-D bidirectionally fails, nodes C and D immediately send fault notification (FN) messages to inform the other nodes of that failure. Subsequently, nodes A, B, and E perform an FDB flush operation. However, the FN message with the highest priority is propagated faster than data frames buffered before that failure in node C. Therefore, nodes A and B may build a defective FDB entry regarding client EX. If nodes A and B transmit data frames to client EX, these messages will be forwarded in the wrong direction and filtered in node C. As client EX does not transmit a message, such an undesirable circumstance lasts until the entry expires.

Case 2. Regardless of a congested condition, inconsistent

FDB entries can also appear in a case in which a failure occurs near the block position. As shown in Fig. 2(b), with a large Ethernet ring network, we assume client DX transmits a data frame to client CX. When link D-E is broken, nodes D and E propagate the FN message as quickly as possible. Then, nodes A, B, and C flush their FDB tables. However, the data frame before that failure is still forwarded through almost all the ring links. It arrives at nodes A, B, and C after the protection procedures are completed. Note that depending on the protection technologies, the second FN message from the other NAF does not trigger FDB flushing. In this situation, nodes A, B, and C receive erroneous data regarding client DX's MAC address, as it reflects the wrong direction.

We introduce just two cases in a single ring, but the FDB inconsistency problem can also occur in other scenarios, such as in multiple interconnected rings. Once the FDB inconsistency problem occurs, it makes ring networks unstable and causes unexpected behavior. Therefore, more suitable protection schemes should be considered.

2. Related Works

The earlier version of this work proposed methods to resolve the FDB inconsistency problem [15]. The set prio scheme was also modeled theoretically and analyzed with numerical simulations. However, the other two methods were simply proposed and described in a qualitative manner. In this paper, quantitative analyses of those methods are supplemented using extensive simulations, and further comparisons are discussed. This paper also observes the stability problem of the flush-based self-healing rings that can occur even in noncongested condition.

Regarding one of our proposed schemes, set prio, there was an initial study on a stability issue with communication protocols. In [16], Shaikh and others studied the stability and robustness of two common Internet routing protocols, namely, open shortest path first (OSPF) and border gateway protocol. To analyze the behavior of routing protocols, they developed an absorbing Markov chain model for estimating the route flap time and the adjacency recovery time. From their models, they determined that the stability of the routing protocols was influenced by the traffic overloads. The analytic model of our set prio scheme is inspired by their approach, but their study focused on a simple system with two or three nodes, whereas we provide a more general analysis concerning input traffic and queue models in an Ethernet ring network with k nodes. Furthermore, we propose an enhanced scheme to successfully distribute the control messages, even in a congested ring. The research conducted by Basu and Riecke in [17] was about the stability of OSPF routing and addressed three issues: the effect of the traffic engineering extensions to the stability of OSPF, subsecond Hello timers to speed up the convergence time, and the synchronization of the link state advertisement (LSA) refresh timers. Through simulations, they showed that the subsecond Hello message helps the OSPF routing protocol to achieve a fast convergence time.

These previous studies analyzed how seriously the congested networks affect the stability of communication protocols. Meanwhile, [18]-[20] focused on reducing the flush operation as much as possible since an FDB flush often introduces a large amount of transient traffic caused by flooded data frames immediately after protection switching. However, we find that such flush optimization can instigate the FDB inconsistency. In section V, we discuss this observation further.

IV. Faultless Protection Schemes to Guarantee Consistent Forwarding Information

In this section, we propose four schemes to guarantee the consistency of an FDB: dual flush, f-timer, purge trigger, and set prio. The dual flush operation is only available for a noncongested ring, whereas the other three schemes remove the erroneous FDB entries in congested rings. In the following subsections, we describe these mechanisms in detail.

1. Dual Flush Operation

Inconsistent FDB entries generated from delayed data frames being transmitted through long ring spans like the case in Fig. 2(b) can be removed by the duplicate flush operations. Normally, ring nodes excluding the NAF receive the FN message from both ring ports in a failure event. Using this fact, the FDB flush operation is individually invoked for each ring port. In other words, each ring node refreshes its FDB at least twice. For example, the faulty FDB entry for client DX at nodes A, B, and C in Fig. 2(b) is eliminated by the subsequent FN message propagated from node E.

2. Flush Delay Timer Setting Scheme

When ring nodes receive the FN message regarding a single failure, an f-timer lets ring nodes prolong an FDB flush operation during its timer period. For example, when the ring nodes in Fig. 2(a) receive the FN message, they individually start their f-timer. Subsequently, the FDB flush operation can be prolonged until all data frames buffered before that failure can be forwarded. In addition, node E in Fig. 2(a) should hold its logical block to prevent abnormal FDB entries while the f-timer activates. The value of this timer should be larger than the maximum time in which a data frame traverses to all links of a

ring and less than 50-ms protection switching time. One global value can be assigned to all the nodes or each node can also configure the value independently of other nodes. Since this method only requires an implementation of a timer to delay the FDB flush operation, it should be a simple solution to eliminate the FDB inconsistency problem.

3. Purge Triggering Scheme

The purging scheme purges all the buffered data frames along with the FDB flush operation. In Fig. 2(a), when node C detects a failure, it no sooner performs an FDB flush operation than purges its buffered data frames. The same operation is performed at each subsequent node receiving the FN message. Therefore, nodes A and B no longer build an FDB entry for client EX. From the implementation view, a control process with this scheme in a ring node should support additional functions to remove data frames buffered in all ring ports as well as the switching buffer in the forwarding module.

4. Priority Setting Scheme

The set prio scheme allows such protection messages as the FN message to be the same as the lowest priority of data frames. Therefore, the FN message is always forwarded after the buffered data frames are completely transmitted. Since this method does not permit the protection message to overtake in-flight data frames, it simply avoids generating incorrect FDB entries. However, it can only be implemented in an Ethernet bridge with the strict priority queue or the first-come, first-served queue.

V. Performance Evaluations

In this section, we evaluate the protection switching performance of the proposed four solutions. The experiments carried out focus on two aspects: the performance of our proposed three schemes in a congested ring and the performance comparison between a single FDB flush and dual FDB flush in a noncongested ring. The performances are simulated and measured by the OPNET simulator [21]. In the former, the three methods, that is, f-timer, purging, and set prio, are plugged into a practical protection switching technology, ITU-T G.8032 ERP. The original standard method (std) is also evaluated to verify how well each scheme eliminates invalid FDB entries in the event of a link failure.

Figure 3 illustrates the simulated ring network in which sixteen ring nodes, M_1 to M_{16}, are individually connected to two adjacent ring nodes through a 20-km 1-Gbps full-duplex link. With a 10-km 1-Gbps link, each node is also linked to one

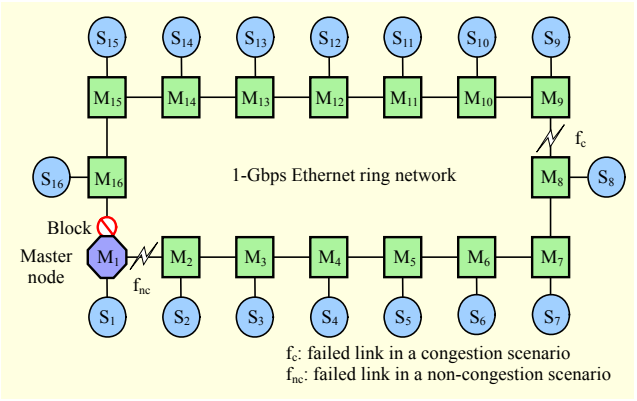

Fig. 3. Simualtion scenarios.

client subnetwork (S_x, $1 \le x \le 16$) where 1,000 clients reside. We suppose a link failure is detected within hundreds of microseconds from an underlying physical layer. The failure detection delay is set to 100 μs in our experiments. We also assume that the ring nodes process data and control messages separately. Referring to a commercial metro Ethernet switch product, the service rates of data and control messages are set to 6.5 mpps (million packets per second) and 0.5 mpps, respectively. Every node schedules Ethernet frames based on their priorities. Each priority queue accommodates 1,000 messages at most. In the following subsections, we lay out the results for each of our experiments in detail.

1. Simulation Results of Congested Ring Networks

The experiment for a congested ring consists of four sets of scenarios. To quantify congestion levels, we develop a traffic intensity formula, which calculates an average traffic amount of all active ring links in each direction. If all traffic generated from clients is equally distributed among all other clients, then the traffic intensity Φ in a ring is simply given by (1), where N, C, k, and λ_i represent the number of ring nodes, the capacity of a link in a ring, the hop count from the NAF (or the node attached to the blocked link), and the data rate from the node that is at i hops away from the NAF in the direction of traffic, respectively. We run evaluations with four different values of Φ,

$$\Phi = \frac{1}{(N-1)\cdot C}\sum_{k=1}^{N-1}\left[\frac{N-k}{N-1}\sum_{i=0}^{k-1}\lambda_i\right], \quad (1)$$

0.5 through 0.8 with a step of 0.1. For each value, each client exponentially generates data frames with the traffic parameters (interarrival time and average frame size), as shown in Table 1. Each scenario is carried out by 10 independent runs.

Having assumed the failure (f_c) happened in link M_8-M_9, we measure the number of incorrect FDB entries, the protection switching time, and normalized data throughput. Referring to

Table 1. Traffic parameters.

Parameter \ Intensity	Noncongested ring				Congested ring			
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8
Interarrival (ms)	140	70	46.5	35	25	22.5	20	17.5
Size (bytes)	600	600	600	600	580	620	620	620

Fig. 4(a), the std always introduces the largest incorrect FDB entries among the others. It is observed that the invalid entries significantly increase in accordance with the traffic intensity. When Φ=0.8, the ratio of incorrect FDB entries is about 3.0%. Meanwhile, the others prevent incorrect entries effectively, but f-timer reveals some defective entries in the biggest intensity, Φ=0.8. This is because the configured timer value is not long enough to process buffered data traffic.

The protection switching time in which all ring nodes complete the protection procedures up to the FDB flush operations is depicted in Fig. 4(b). First, f-timer has the longest protection switching time. Its results show a constant value of about 13 ms because a global timer value of 10 ms is configured in all scenarios. Then, set prio shows that the more that traffic increases, the longer the protection completion time is stretched by the amount of the buffered data. However, the average protection time in the worst case is only about 9 ms less than f-timer. Unlike those methods, purging and std have the lowest protection completion time regardless of congested conditions. The two methods finish the protection procedures within 3 ms.

Finally, we evaluate the total amount of data frames that all clients receive during the first second following the failure. As portrayed in Fig. 4(c), the packet delivery ratios of all methods decline as the traffic intensity increases. Due to inconsistent FDB entries, the std shows the lowest packet delivery ratios at all traffic intensities. The f-timer accepts more data packets than the std but represents a small decrease compared to the number accepted by the set prio and purging schemes. The set prio and purging show comparable packet delivery ratios in all cases, but the result of set prio is slightly higher than that of purging.

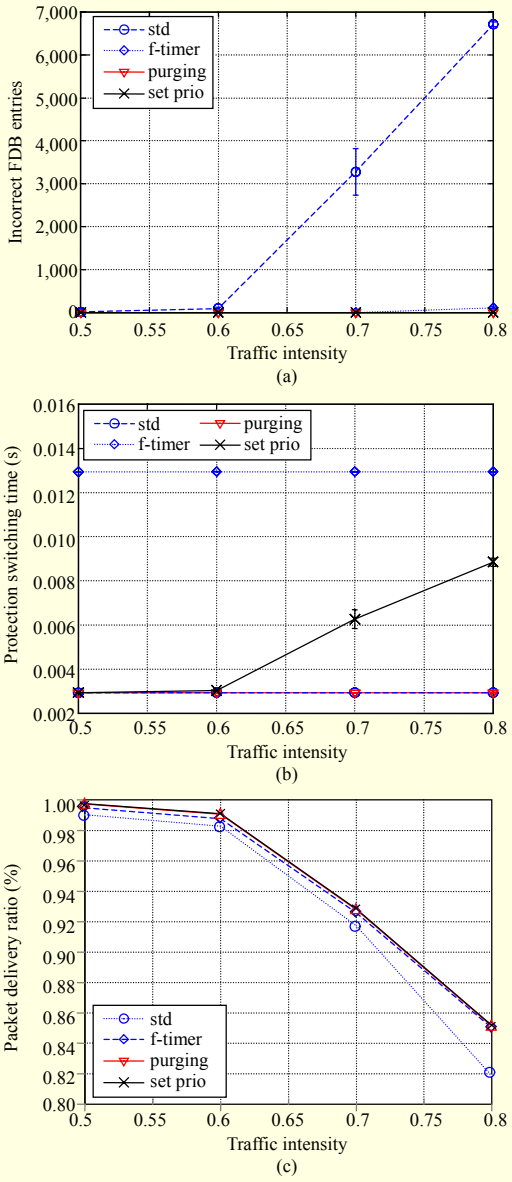

Fig. 4. Comparison of results in congested ring: (a) number of incorrect FDB entries, (b) protection switching time, and (c) packet delivery ratio.

2. Simulation Results of Noncongested Ring Networks

As mentioned in subsection IV.1, inconsistent FDB entries in a noncongested ring can be removed by the dual flush operation. To demonstrate that the dual flush scheme is more stable than a single flush scheme, we observe how many defective FDB entries each method produces. Interestingly, such examination can be carried out with two versions of G.8032's; that is, G.8032v1 [6] and G.8032v2 [7]. Ring nodes with G.8032v1 process its protection activities depending on a node's state so that they perform a single FDB flush operation normally. On the contrary, G.8032v2 performs FDB flush based on the port and node information in the protection message and lets ring nodes execute the flush operation twice. Using these features of G.8032, our experiments are carried out in noncongested rings with four different traffic intensities (0.1 to 0.4 with a step of 0.1). The traffic parameters are specified in Table 1.

The failure event (f_{nc}) is assumed to occur in link M_1-M_2, as shown in Fig. 3. As shown in Fig. 5, incorrect FDB entries are

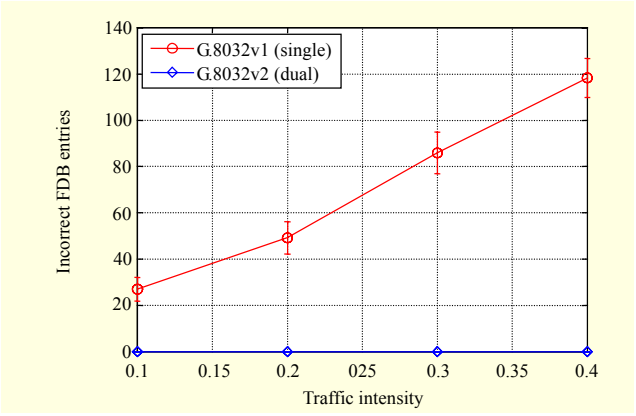

Fig. 5. Number of incorrect FDB entries in noncongested rings.

generated only in the single flush operation. This result demonstrates that the dual flush scheme definitely removes invalid FDB entries. To check if the duplicate flush operation does not produce any more transient traffic than the single flush, we additionally measure the average link utilization during the second following the failure. Although we do not show the result due to the space limitation, the gap between two schemes is less than 1% in every scenario. The gap is minimal because the interval between the arrivals of two protection messages received in both ring ports is rather short in all non-NAF nodes.

3. Discussion

In the congested ring scenarios, the flush timer scheme makes the master node hold its logical block until the flush timer expires. It causes the rings to be segmented during the timer period and leads to loss of data frames. To determine the timer value, the maximum delay that a data frame traverses its ring networks must be known to operators to achieve a successful protection process. The purging method can also cause loss of data frames due to the forced purge operation. This scheme needs to clear data frames buffered in the MAC entity of each ring port as well as the bridge's MAC relay entity. The set prio scheme in the congested rings solves the FDB inconsistency problem without any additional implementation. However, it introduces a prolonged flush operation like the flush timer even though its delay is shorter than the flush timer period in most cases.

In the noncongested scenarios, G.8032v1 allows some nodes to build incorrect FDB entries, but G.8032v2 does not permit them to do so without a distinguishable generation of transient traffic. There have been some proprietary and standard recommended options to reduce the number of flush operations down to a single operation. However, our simulation results imply that such flush optimization methods are rather dangerous and more meticulous flush optimization methods should be considered. We do not consider any interconnected ring network scenarios here, but the single flush operation under a failure event in a subring can produce significantly faulty FDB entries in both the major rings and subrings. The change of subring topology affects forwarding routes of traffic traversing both rings.

In our experiments, we consider an Ethernet ring with sixteen ring nodes, but practical transport networks might have scores of nodes owing to several rings being interconnected. In such large rings, the defective protection can be much more severe. The four presented methods do not affect the original protection procedures and can be easily implemented in network facilities. In addition, we recommend an approach to combine dual flush with one of the three solutions for congested rings for the faultless protection. The combined scheme provides a more flexible solution for different traffic levels on the ring.

VI. Analytic Model of Ethernet Ring Protection Protocol with Priority Setting Scheme

In this section, we develop an analytic model to observe how the protection messages with the lowest priority are affected by congestion. To theoretically estimate the expected protection completion time, an absorbing Markov chain is modeled here. Note that our analytic model is also based on a practical protection method, the ERP protocol. Also, the estimated protection time considers only one ring direction to provide an intuitive understanding of our analytic model.

1. Analytic Model of ERP Protocol

The behavior of the ERP protocol under a failure condition is modeled as an absorbing Markov chain in Fig. 6. This Markov chain also shows a delay value associated with each transition along with the transition probabilities, since delays between states are different. First, all ring nodes remain in the normal state S_1. When a link failure occurs, an NAF directly enters into the protection state S_5 without any delay, since it detects the failure by its own defect logic. Meanwhile, the NAF multicasts an R-APS(SF) message in both ring ports. The ERP protocol lets NAFs multicast a burst of three R-APS(SF) messages to cope with the loss of one or two messages. The burst interval of each R-APS(SF) message, α, is normally set to 3.33 ms. After the first three messages are transmitted, one R-APS(SF) message continues to transmit at a regular interval, β, which is normally set to 5 s. Accordingly, if a ring node that is n hops away from the NAF receives the message at once, it reaches the protection state S_5, with a $1-\pi_n$ probability and a τ_n message propagation delay. Otherwise, the node reaches the other

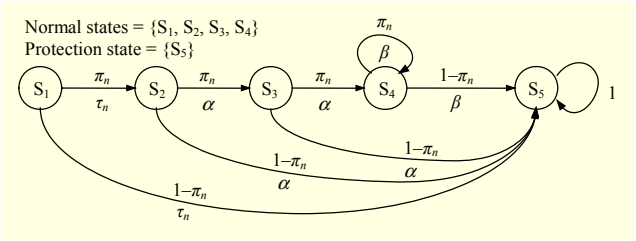

Fig. 6. Absorbing Markov chain of ERP protocol under single failure.

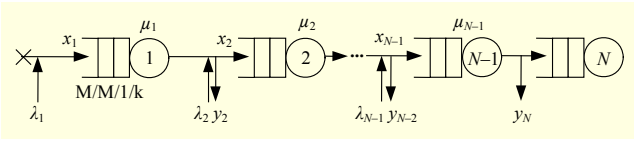

Fig. 7. Ethernet ring with ring nodes having M/M/1/k queue.

normal state S_2, with a π_n loss probability and a τ_n propagation delay. In this context, if the node does not receive any of the burst messages, it reaches the normal state S_4, with the accumulated values of both the propagation delays and the burst intervals. As a result, if the ring node also fails to receive the message in that state, it remains in the normal state S_4, with the probability of π_n and the interval of β. This state persists while the node continues to miss the SF messages.

Suppose an Ethernet ring consists of N ring nodes; then, the propagation delay τ_n in a node that is n hops away from an NAF is computed using the following equation:

$$\tau_n = \sum_{i=0}^{n-1}\left[\delta_i + \omega_i\right], \quad \forall n, 1 \le n \le N-1, \tag{2}$$

where δ_i is the time during which the R-APS(SF) message is sojourned at a node that is i hops away from the NAF, and ω_i is the propagation delay associated with the link connecting the i-th node and the $(i+1)$th node. Meanwhile, the loss probability of R-APS(SF) in a node that is n hops away from an NAF π_n is calculated by the following equation:

$$\pi_n = 1 - \sum_{i=0}^{n-1}(1-\sigma), \quad \forall n, 1 \le n \le N-1, \tag{3}$$

where σ_i is the loss probability of the R-APS(SF) message at an output queue of the i-th node.

Based on the two equations above, all elements $p_{n,i,j}$ and $t_{n,i,j}$ of matrices $\mathbf{P}_n$ and $\mathbf{T}_n$ to represent the transition probabilities and the transition delays of a node that is n hops away from an NAF can be obtained. Then, using the theory of absorbing the Markov chain, the fundamental matrix $\mathbf{F}_n$, which describes the expected number of times between two nonabsorption states, can be obtained by the inverse of matrix $\mathbf{I}$-$\mathbf{Q}_n$, where $\mathbf{I}$ is the identity matrix and matrix $\mathbf{Q}_n$ represents all the transient movements except the absorption states in matrix $\mathbf{P}_n$. With all elements $f_{n,i,j}$, $p_{n,i,j}$, and $t_{n,i,j}$ in $\mathbf{F}_n$, $\mathbf{P}_n$, and $\mathbf{T}_n$, the expected protection switching time from the starting state S_1 to the terminating state S_5 in a node that is n hops away from NAF, ε_n, can be obtained by the following equation:

$$\varepsilon_n = \sum_{j=1}^{4}\left(f_{n,1,j}\left[\sum_{k=1}^{5} t_{n,j,k}\, p_{n,j,k}\right]\right) = \tau_n + \alpha\pi_n + \alpha\pi_n^2 + \frac{\beta\pi_n^3}{1-\pi_n}. \tag{4}$$

Consequently, given both the burst and regular intervals for R-APS(SF), if the loss probability of the SF message, π_n, and the propagation delay, τ_n, are obtained in each node, the expected protection switching time in each node can be solved easily by (4). To obtain π_n and τ_n for ring nodes, we model an Ethernet ring with the M/M/1/k queue system in the following subsection.

2. Queue Model of Ethernet Ring Network

An Ethernet ring underlying the M/M/1/k queue is portrayed in Fig. 7. The Ethernet ring consists of N ring nodes with three ports: east, west, and subnet. The M/M/1/k queue has exponential interarrival time and service time distributions, each with the respective parameters λ and μ. This queue characterizes that it is stable even for $\rho=\lambda/\mu>1$. Let us assume node 1 is the RPL owner (master node). Supposing each subnet exponentially generates the amount of traffic toward destinations equally distributed among all other subnets in the ring network, the effective arrival rate, not including loss of arrivals in a node that is k hops away from NAF, is given by

$$x_k = \frac{N-k}{N-1}\sum_{j=1}^{k}\left[\lambda_j \prod_{i=j}^{k}(1-\sigma_i)\right], \tag{5}$$

where $1 \le k \le N-1$, λ_i is the arrival rate from the subnet of node i, and σ_i is the loss probability of node i. The rate of the traffic being dropped from the ring and delivered to the subnet port of node i, y_i, is taken into account for calculating the value of x_k. Subsequently, using Little's formula, the loss probability and average sojourn time with respect to the output queue of each node can be easily obtained.

3. Evaluation of Expected Protection Time

Considering the analytic model above, we analyze the lowest bound of the expected protection switching time in each node under a single link failure. The scenario is modeled as the Ethernet ring that consists of five ring nodes wherein each node has a subnet that exponentially generates the amount of traffic toward destinations equally distributed among all other subnets in the Ethernet ring. The fault condition of each ring link is assumed to be detected from the underlying physical layer with a negligible delay. It is also assumed that the failure event occurs in a blocked link. When this link fails, the ring nodes do

not need to perform the FDB flush operation because the ring topology is unchanged. Since no transient traffic is generated, the lowest bound of the expected protection completion time can be computed. To evaluate how the congested condition affects the stability of the ERP protocol with the lowest priority R-APS(SF) message, we estimate the expected protection completion time, focusing on the following two aspects: different ring capacities and different traffic overloads. The evaluation results follow.

A. Expected Protection Time versus Ring Capacity

With the developed analytic model, the protection switching performance is also evaluated using the NS-2 simulator [22] to validate the mathematical approach. Based on the modeled scenario above, we separately evaluate the expected protection time in both 100-Mbps and 1-Gbps Ethernet rings. Each subnet exponentially transmits 80-Mbps and 800-Mbps traffic with an average of 512 bits per message to generate traffic congestion corresponding to Φ=1. The output buffer of every node in both rings can accommodate 1,000 messages. The simulation for each ring is carried out by 1,000 independent runs. Note that our evaluations show only results with respect to one ring direction. We number the nodes from 0 to 4, as shown in Fig. 8. For instance, the RPL designated as the NAF is node 0.

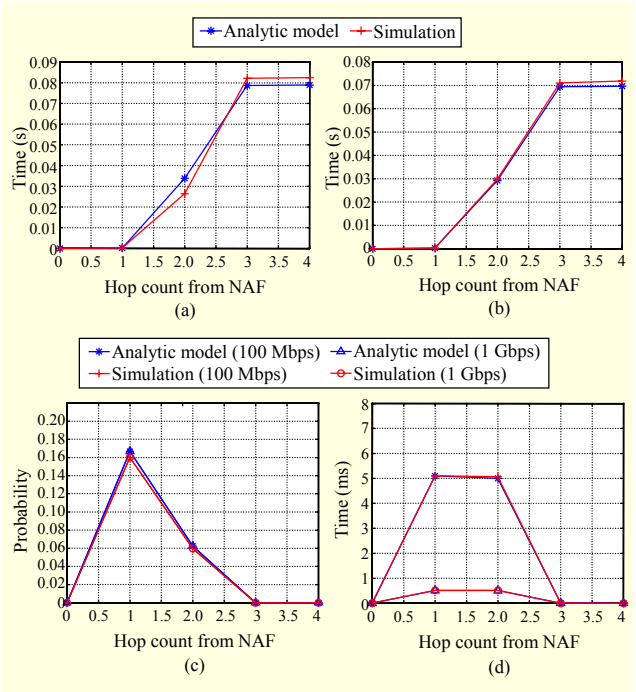

Fig. 8. Comparison of results between simulations and analytic model: expected protection time (a) in 100-Mbps ring, (b) in 1-Gbps ring, (c) loss probabilities, and (d) mean sojourn time in both rings.

As shown in Fig. 8(c), both rings have almost the same loss probabilities due to the same traffic intensity. Congestion is observed in nodes 1 and 2. As shown in Figs. 8(a) and 8(b), the expected protection times obtained from the simulation and the analytic model are almost the same, and these results show that our analytic model is accurate. Meanwhile, the 100-Mbps ring yields a protection time of about 80 ms, which is more than 1.5 times the 50-ms protection switching time requirement. On the other hand, the 1-Gbps ring finishes the protection switching within about 70 ms following the failure. This value is less than the outcome from the 100-Mbps ring because the average sojourn time in each node on the 1-Gbps ring is shorter than that of the 100-Mbps ring. These results are depicted in Fig. 8(d). As the expected protection time of the last node, that is, node 4, is longer than 50 ms in both cases, it is clearly shown that the congestion affects the expected protection time. Since we set the example of an Ethernet ring with just five nodes, its delay may not be significant in terms of the 50-ms protection switching time requirement. However, if the number of nodes increases or the congestion conditions worsen, we can then assume that the expected protection completion time will be significantly prolonged.

B. Expected Protection Time versus Traffic Intensity

To observe how the amount of traffic influences the protection switching time, we simulate a 100-Mbps Ethernet ring using the analytic model. As shown in Fig. 9(a), the most congested condition (Φ=1.2) yields the longest expected protection time of about 500 ms. This result is as much as ten times the desired 50-ms protection time. As shown in Fig. 9(b), the same congested condition has a message drop probability of up to 33%. However, we see that the congested situation of Φ=0.9 produces a loss probability of less than 8% and a protection time of less than 10 ms. Meanwhile, the lowest traffic intensity (Φ=0.8) satisfies the sub-50-ms protection switching time requirement without any loss probability.

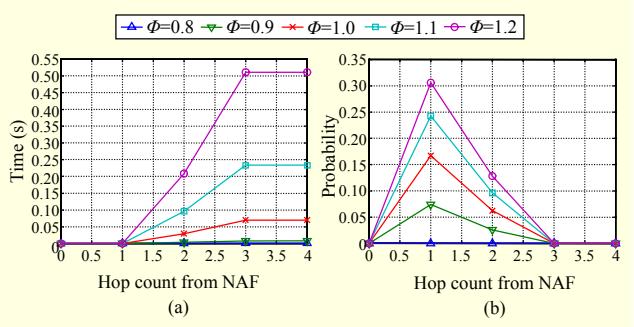

Fig. 9. Analytic results with different traffic intensities: (a) expected protection times and (b) loss probabilities.

VII. Enhanced Priority Setting Scheme

The results in the previous section imply that the priority scheme does not perform to meet the protection switching time requirement in a congestion condition due to the loss of the protection messages. To resolve such a drawback, we suggest using an enhanced set prio method. The basic idea is to find an appropriate number for the burst transmission of the failure notification message in accordance with the congestion level of the ring. In some self-healing rings (not in ERP), the master node multicasts a control message periodically to detect the health of the ring in the normal state. If the health check message initiated from a ring port of the master node is returned to the opposite ring port, the ring is consequently regarded as normal. Otherwise, the master node waits for a specified number of times, after which it considers the link as broken and then can start the protection procedures.

By adopting a type of control message that circles around the ring, the master node is able to monitor the round-trip time (RTT) of that message per each round and then determine the current congestion level of its ring. Depending on the measured RTT values, which indicate the level of congestion on the ring, the master node decides the appropriate number for the burst transmission, which is called "the burst number (γ)." An algorithm to compute the burst number based on the measured RTT values is proposed in [15], and other variants are possible. Finally, the master node conveys the current burst number information in the next health check message so that other ring nodes can also know the same burst number that the master node knows. In turn, all the ring nodes are able to prepare an optimal burst number to guarantee successful delivery of protection messages in a future failure event.

To examine how much the enhanced scheme can improve the protection completion time in congested conditions, we evaluate the protection completion time in ERP according to various burst numbers (γ). Normally, an RPL owner transmits one R-APS(no request) message every 5.0 s. Regarding the two rings (1-Gbps and 100-Mbps) described in section VI, the evaluated values are depicted in Figs. 10(a) and 10(b). In both rings, a one-step growth from the default size (γ=3) decreases the protection completion time more than 50 ms. The larger the burst number is, the faster the protection performs. In a general network with a mesh topology, congestion information obtained by such an RTT measure cannot be reliably used to predict packet loss because of the dynamics of the routes. It also requires additional bandwidth for detecting the network status. However, the path on an Ethernet ring is unique and fixed, and the health check flows do not consume any significant percentage of the total traffic. Therefore, our enhanced scheme is feasible to implement to help ring nodes suffering from a failure to perform the reliable protection switching.

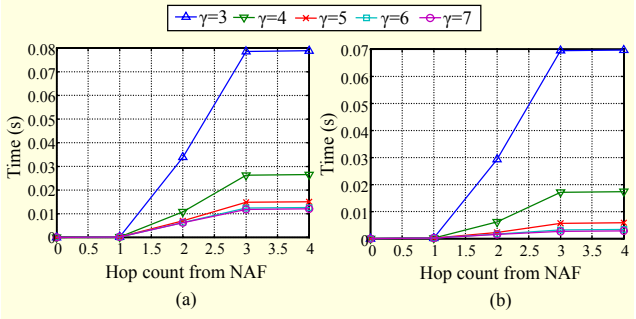

Fig. 10. Analytic results in various burst numbers: expected protection time (a) in 100 Mbps ring, and (b) in 1 Gbps.

VIII. Conclusion

This paper discussed a defective protection method of self-healing Ethernet rings that delays data frames generated before a failure occurs, causing ring nodes to generate erroneous forwarding information. To resolve the undesirable situation, we proposed four simple methods: dual flush, f-timer, purging, and set prio. The experiment results in both noncongested and congested rings demonstrated that these methods effectively suppress inconsistent FDB entries. The enhanced version of the set prio, which emerged from the observations of our analytic model, also achieved fast and reliable protection, even in a congested condition. All of these solutions can be easily implemented in the original protection procedures without increasing the complexity. We believe that they guarantee faultless protection in self-healing Ethernet ring networks.

References

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

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

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

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

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

[6] Ethernet Ring Protection Switching, ITU-T Rec. G8032/Y.1344 (Ver 1.0), 2008.

[7] Ethernet Ring Protection Switching, ITU-T Rec. G8032/Y.1344

(Ver 2.0), 2010.
[8] J. Ryoo et al., "Ethernet Ring Protection for Carrier Ethernet Networks," *IEEE Comm. Mag.*, vol. 46, no. 9, Sept. 2008, pp. 136-143.
[9] Extreme Networks' Ethernet Automatic Protection Switching (EAPS) Version 1, IETF RFC 3619, 2003.
[10] Huawei Corp., "Technical White Paper for Rapid Ring Protection Protocol (RRPP)," Huawei Corp., Tech. Rep., 2007.
[11] Cisco Systems, "White Paper for Resilient Ethernet Protocol (REP)," Cisco Systems, Tech. Rep., C11-427224-00, 2007.
[12] IEEE Standard for Local and Metropolitan Area Networks – Virtual Bridged Local Area Networks, Amendment 5: Connectivity Fault Management, IEEE Std. 802.1ag, 2007.
[13] OAM Functions and Mechanisms for Ethernet Based Networks, ITU-T Rec. Y.1731, 2006.
[14] J. Ryoo et al., "OAM and Its Performance Monitoring Mechanisms for Carrier Ethernet Transport Networks," *IEEE Commun. Mag.*, vol. 46, no. 3, pp. 97-103, Mar. 2008.
[15] K. Lee, C. Lee, and J. Ryoo, "Enhanced Protection Schemes to Guarantee Consistent Filtering Database in Ethernet Rings," *Proc. IEEE Global Comm. Conf. (Globecom)*, Dec. 2010.
[16] A. Shaikh et al., "Routing Stability in Congested Networks: Experimentation and Analysis," *Comput. Comm. Rev.*, vol. 30, no. 4, 2000, pp. 163-174.
[17] A. Basu and J.G. Riecke, "Stability Issues in OSPF Routing," *Comput. Comm. Rev.*, vol. 31, no. 4, 2001, pp. 225-236.
[18] K. Lee and J. Ryoo, "Flush Optimizations to Guarantee Less Transient Traffic in Ethernet Ring Protection," *ETRI J.*, vol. 32, no. 2, Apr. 2010, pp. 184-194.
[19] 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.
[20] D. Lee et al., "Efficient Ethernet Multi-ring Protection System," *Proc. 7th IEEE Int. Workshop Design Reliable Commun. Netw. (DRCN)*, 2009, pp. 305-311.
[21] OPNET Technologies Inc. Available: http://www.opnet.com
[22] The Network Simulator – NS2, 1995. Available: http://www.isi.edu/nsnam/ns

Kwang-koog Lee received his BS and MS in electronics and communication engineering from Kangwon National University, Chuncheon, Rep. of Korea, in 2006 and 2008. He received his PhD in broadband convergence network engineering from the University of Science and Technology (UST), Daejeon, Rep. of Korea, in 2012. From 2008 to 2012, he worked at ETRI, Daejeon, Rep. of Korea, where he studied switching and management technologies for high-speed optical transmission systems. He is currently working as a postdoctoral research fellow at the Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Rep. of Korea. His research interests include packet-based transport networks, such as Carrier Ethernet and MPLS-TP, optimized protocol design and performance analysis in wired/wireless networks, and the Internet of Things.

Jeong-dong Ryoo is a principal researcher at ETRI, Daejeon, Rep. of Korea. He holds an MS and a PhD in EE from the Polytechnic Institute of NYU, NY, New York, USA, and a BS in EE from Kyungpook National University, Daegu, Rep. of Korea. After completing his PhD in the area of telecommunication networks and optimization, he started working for Bell Labs, Lucent Technologies, Holmdel, NJ, USA, 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 joined ETRI in 2004, his work has been focused on next-generation networks, Carrier Ethernet, and MPLS-TP 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 the Eta Kappa Nu association.

Bheom Soon Joo received his BS in electronics engineering from Seoul National University, Seoul, Rep. of Korea, and his MSc in electronics engineering from the Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Rep. of Korea. He has worked with ETRI since 1984, where he is currently a principal engineer and the director of the Optical Packet Integrated Research Team. His research interests span the fields of development for Carrier Ethernet switch systems, hardware development ATM switch systems, network synchronization, and high-speed interconnection.