Protection Switching Methods for Point-to-Multipoint Connections in Packet Transport Networks

Dae-Ub Kim, Jeong-dong Ryoo, Jong Hyun Lee, Byung Chul Kim, and Jae Yong Lee

In this paper, we discuss the issues of providing protection for point-to-multipoint connections in both Ethernet and MPLS-TP-based packet transport networks. We introduce two types of per-leaf protection–linear and ring. Neither of the two types requires that modifications to existing standards be made. Their performances can be improved by a collective signal fail mechanism proposed in this paper. In addition, two schemes — tree protection and hybrid protection — are newly proposed to reduce the service recovery time when a single failure leads to multiple signal fail events, which in turn places a significant amount of processing burden upon a root node. The behavior of the tree protection protocol is designed with minimal modifications to existing standards. The hybrid protection scheme is devised to maximize the benefits of per-leaf protection and tree protection. To observe how well each scheme achieves an efficient traffic recovery, we evaluate their performances using a test bed as well as computer simulation based on the formulae found in this paper.

Keywords: Protection switching, packet transport network, point-to-multipoint, P2MP.

Manuscript received Feb. 6, 2015; revised Oct. 7, 2015; accepted Oct. 21, 2015.

This work was supported by the ICT R&D program of MSIP/IITP (No. B0101-15-0024, Terabit optical-circuit-packet converged switching system technology development for the next-generation optical transport network).

Dae-Ub Kim (artkdu@etri.re.kr) is with the Communications & Internet Research Laboratory, ETRI, and with the Department of Information and Communications Engineering, Chungnam National University, Daejeon, Rep. of Korea.

Jeong-dong Ryoo (corresponding author, ryoo@etri.re.kr) is with the Communications & Internet Research Laboratory, ETRI, Daejeon, and with the Department of Engineering, Korea University of Science and Technology, Daejeon, Rep. of Korea.

Jong Hyun Lee (jlee@etri.re.kr) is with the Communications & Internet Research Laboratory, ETRI, Daejeon, Rep. of Korea.

Byung Chul Kim (byckim@cnu.ac.kr) and Jae Yong Lee (jyl@cnu.ac.kr) are with the Department of Information and Communications Engineering, Chungnam National University, Daejeon, Rep. of Korea.

I. Introduction

Packet transport network (PTN) technologies, such as Transport Ethernet and Multiprotocol Label Switching – Transport Profile (MPLS-TP), are rapidly gaining in importance as they become main solutions in the area of transport networks, which has traditionally been based on synchronous digital hierarchy (SDH) and optical transport networks (OTNs). The boundaries between packet and circuit networks are disappearing as many traditional circuit-switched applications such as voice and video are now being carried over packet-switched MPLS or Ethernet networks.

Recently, many efforts have progressed to develop operations, administration, and maintenance and protection switching, which constitute key technologies for promoting Ethernet and MPLS into transport networks. Such activities have mainly been conducted in the International Telecommunication Union-Telecommunication Standardization Sector (ITU-T) Study Group 15 in close collaboration with the Institute of Electrical and Electronics Engineers and the Internet Engineering Task Force (IETF).

Linear protection switching provides active measures to react against cable cuts, node failures, or signal degradation on an end-to-end connection, which can be seen as an Ethernet Virtual Local Area Network (VLAN) or MPLS label-switched path (LSP). For Ethernet linear protection, the Automatic Protection Switching (APS) protocol is specified in the ITU-T Recommendation G.8031 [1], whereas the Automatic Protection Coordination (APC) protocol is specified in both the IETF RFC 7271 [2] and the ITU-T Recommendation G.8131 [3] for MPLS-TP. The APC protocol was created to enhance the Protection State Coordination (PSC) protocol specified in the IETF RFC 6378 [4] and to provide operator control and

http://dx.doi.org/10.4218/etrij.16.0115.0024

experience that more closely models the behavior of linear protection seen in the APS protocol. A detailed description of the APC protocol and its relationship with the APS and PSC protocols can be found in [5].

Ethernet ring protection (ERP), which has been developed as part of the ITU-T Recommendation G.8032 [6], can support both point-to-point (P2P) and point-to-multipoint (P2MP) services, but it is optimized for a ring topology utilizing generic mechanisms inherited from the traditional Ethernet frame header and bridge functions [7]. It is also worth mentioning that ERP is well known for having issues with the Filtering Database (FDB) flush operation, which causes all ring nodes to broadcast data frames until source address learning is complete. To obtain stable protection switching performance, some schemes solving network stability issues related to FDB flush and ERP have been discussed in [8]–[12].

With the popularity of video distribution, IPTV, and other one-to-many applications, both service providers and network operators need their premium P2MP services to be protected. A wide variety of P2MP services can be efficiently realized over P2MP connections in Ethernet and MPLS-TP-based PTN. To provide resiliency for P2MP connection networks, protection switching capability, as seen in other topologies, should be provided to recover traffic within the typical transport network protection time requirement (that is, below 50 ms) in the event of a network defect.

This paper focuses on protection switching methods for a P2MP connection that is transported over an Ethernet or MPLS-TP network. The protection switching schemes are designed to provide protection over a multipoint service, which is commonly referred to as an Ethernet-Tree service [13], connecting one root and a set of leaves, but preventing the leaves from communicating directly without passing through the root.

In this paper, we propose four schemes to provide protection switching capabilities for P2MP connection networks — per-leaf protection with existing linear protection, per-leaf protection with existing ring protection, tree protection, and a hybrid of per-leaf and tree protections. While two per-leaf protections can be used without any modification to the existing protection standards, their performances can be improved by a collective signal fail (C-SF) mechanism — one that is to be proposed in this paper. Tree protection is proposed to reduce the service recovery time when a single failure leads to multiple signal fail (SF) events and results in a root node having to bear a significant processing burden. A hybrid of per-leaf and tree protections is also newly proposed to maximize the benefits of per-leaf and tree protections. To observe how well each scheme achieves efficient protection in a failure event, we evaluate their performance using a test bed with real systems and computer simulations based on mathematical formulae.

The remainder of this paper is divided into three sections. Section II presents a reference network model and protection switching schemes for resilient P2MP networks. Performances of the presented schemes are evaluated in Section III, and finally, conclusions are drawn in Section IV.

II. Reference Network Model and Protection Switching Schemes for Resilient P2MP Networks

A reference network for a resilient P2MP service network can be modelled as shown in Fig. 1. A root node, *R*, and a set of leaf nodes (*L*1, … , *L*9) are connected over a logical tree structure, which may branch out at intermediate nodes (*I*1, … , *I*4). A protection tree (dashed lines) provides protection against a failure in a working tree (solid lines), which is used for traffic delivery in a fault-free situation. The two trees are completely disjointed to prevent a single point of failure and pre-constructed prior to a failure for fast traffic recovery.

In the following subsections, we propose four protection switching schemes for P2MP networks. The main design objective of the proposals is to reuse any existing protection switching technologies as much as possible.

1. Per-leaf Protection with Existing Linear Protection

A P2MP network can be protected with multiple P2P linear protections. As shown in Fig. 2, one P2P working path and one P2P protection path are prepared for each leaf node. For example, the working path between the root node and *L*1 leaf node, *R*-*I*1-*I*3-*L*1, is backed by the protection path *R*-*I*2-*I*4-*L*1. The number of linear protection processes (LPPs) pertaining to the root node is equal to the number of leaf nodes, whereas that for a leaf node is equal to only one. A pair of LPPs — one belonging to a leaf node and one to the root node — provide P2P linear protection to the corresponding working and

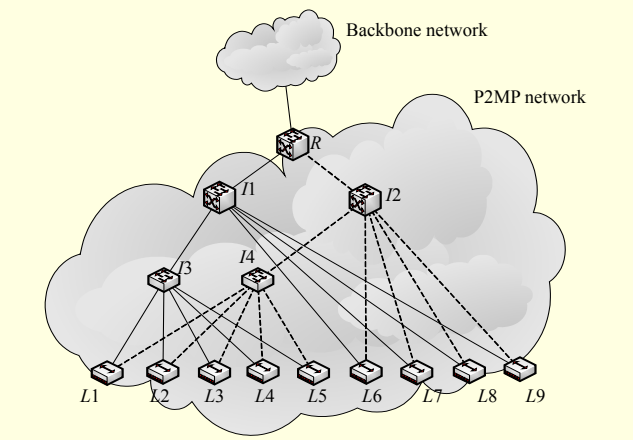

Fig. 1. Resilient P2MP service network.

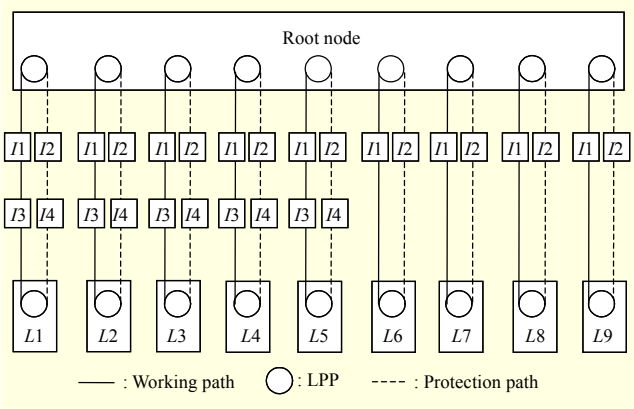

Fig. 2. Per-leaf protection scheme.

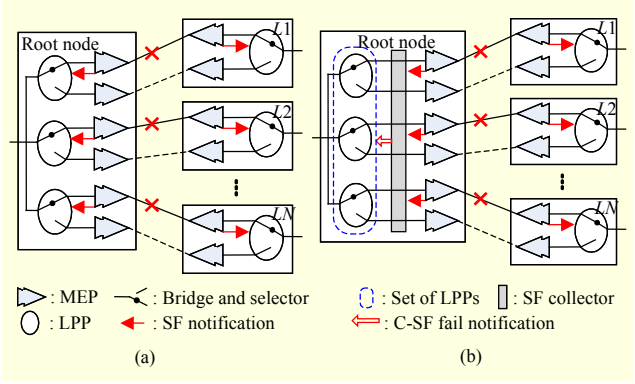

Fig. 3. Internal operations of per-leaf protection: (a) native mechanism and (b) C-SF mechanism.

protection paths and may be chosen independently of other existing LPP pairs. When a defect occurs at any link or node along a working path, the corresponding protection path is used to deliver traffic. We call this scheme per-leaf protection in this paper.

In per-leaf protection, the existing Ethernet and MPLS-TP linear protections can be used without the need for modification to Ethernet and MPLS-TP environments, respectively. Some operation examples of Ethernet linear protection and MPLS-TP linear protection are illustrated in the appendices of ITU-T Recommendation G.8031 [1] and IETF RFC 7271 [2], respectively.

Considering the fact that both Ethernet and MPLS-TP linear protection technologies can provide a sub–50 ms protection switching time regardless of the number of intermediate nodes between two end nodes as long as their distance is less than 1,200 km, per-leaf protection can also provide carrier-grade resiliency for P2MP networks without any modification to the existing standardized mechanisms. However, depending on the location of failure in a network, per-leaf protection can suffer a performance issue. When a cable-cut or a node failure occurs near leaf nodes and the number of affected paths is rather small, all LPPs can complete their protection switching operation within the required protection switching time. On the other hand, if the problem occurs near the root node, then any traffic recovery may be delayed due to the signaling associated with excessive simultaneous SF triggers.

Figure 3(a) shows detailed operations inside the root and leaf nodes when a problem occurs at the link between the root node and intermediate node *I*1. For the sake of brevity, intermediate nodes are omitted from the figure. To monitor the continuity of a path between the root node and a leaf node, a pair of Maintenance Entity Group End Points (MEPs) is activated at the end of each path [14].

If an MEP detects, via a continuity check (CC) function, an anomaly that results in a loss of continuity (LOC) defect, then it informs an LPP of an SF condition that has been detected. Then, the LPP runs its protection switching algorithm to determine the position of the selector and bridge to switch over traffic, and generates APS or APC protocol messages to coordinate the switching action with the other end.

The SF notification event is normally executed by an interrupt-driven algorithm and has higher priority than general periodic tasks in implementations. If a defect occurs along the *R*-*I*1 link in Fig. 1 that simultaneously affects all of the paths that constitute the working tree, then all of the LPPs need to be notified of such an SF. The subsequent mass of sequential operations that follows increases node processing burdens and delays traffic recovery in implementations. In particular, multiple inter-processor communications (IPCs) between MEPs and LPPs in different line-card processors negatively affect the recovery performance.

To relieve the burden related to IPC for SF notifications, all SFs occurring within a certain time interval are collected into a single notification and delivered to the place where all of the LPPs reside. As shown in Fig. 3(b), an SF collector performs the C-SF mechanism. After the C-SF notification arrives at the set of LPPs, each of the collective SFs within the single notification is processed individually. In Section III, the performance benefits of the C-SF mechanism are evaluated.

2. Per-leaf Protection with Existing Ring Protection

Another possible way to protect a P2MP connection is to set up multiple P2P connections with multiple ring protection algorithms, as shown in Fig. 4. In Fig. 4, we assume that nine ERP instances are used for nine rings; the first five rings consist of six nodes (*R*-*I*1-*I*3-*Li*-*I*4-*I*2) and the other four rings consist of four nodes (*R*-*I*1-*Li*-*I*2), where *Li* is leaf node *i* on each ring.

Any one of the nodes on a ring can be configured as the Ring Protection Link (RPL) owner node. In Fig. 4, the root node is assigned to be the RPL owner and the *I*2 node is configured as

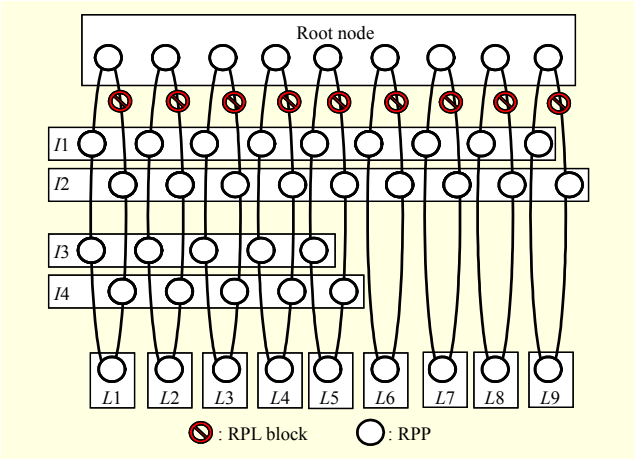

Fig. 4. ERP application for per-leaf protection.

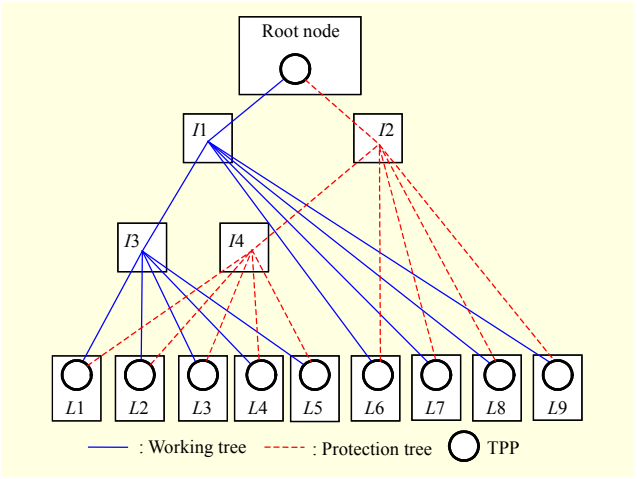

Fig. 5. Tree protection scheme.

the RPL neighbor node. Then, the link *R*-*I*2 is logically blocked to prevent a loop on the ring under normal operating conditions. When a ring node detects an SF condition, which occurs at the link directly connected to its ports, the ring node blocks the traffic on the failed ring port and transmits Ring-APS messages to indicate the presence of the SF condition. Upon receiving the SF messages, the RPL owner unblocks the RPL block. The RPL neighbor also unblocks the RPL block if it is configured to place the RPL block in a normal operating condition. Detailed operations of ERP can be found in [7].

In a manner similar to that of the linear protection application for per-leaf protection, this method requires that multicast service packets be replicated and sent to multiple P2P connections at the root node; moreover, the method has very poor scaling characteristics and consumes vast amounts of bandwidth resources.

3. Tree Protection

In tree protection, a dedicated-protection rooted multipoint connection (protection tree) is prepared to back up a working rooted multipoint connection (working tree). Any failure on a working tree, even if the failure affects only a portion of the leaf nodes, will cause all traffic flowing on the working tree to be switched to the protection tree.

As shown in Fig. 5, tree protection requires a single tree protection process (TPP) in the root node and one TPP in each leaf node. The bridge and selector in a root node and those in each leaf node are coordinated by a protection protocol to hold the same bridge and selector positions. Protection protocol messages should be communicated between a root node and each leaf node to coordinate their bridge and selector positions. As in all the existing protection protocols, the protection protocol messages are necessary to deliver the network operator's commands and the SF detected by an end; that is, unidirectional failure.

A multicast protection protocol message generated by the root node is delivered to all the leaf nodes, while the protocol messages generated by each leaf node are delivered only to the root node. The protection protocol messages from each leaf node should be identifiable to the root node. This can be done by assigning different connection IDs, such as VLAN IDs for Ethernet and LSP IDs for MPLS-TP, or by including a node ID from which the root node can identify the source of the protection protocol message. If a root node initiates protection switching, then tree protection will be completed by a single message exchange between the root node and all the leaf nodes, and the existing single-phase linear protection protocols can be used as is. However, if the protection switching is initiated by a leaf node, then it requires two message exchanges to complete tree protection switching. In the following subsections, we show how the existing Ethernet linear protection protocol can be modified to support tree protection in cases of SFs on the working tree. Although we restrict the scenarios in this paper due to the page limit, more complex scenarios and a complete state machine can be devised based on the following scenarios. In MPLS-TP environments, the APC protocol can be modified similarly.

A. Case of Detecting SFs at Leaf Nodes

Figure 6(a) shows an example scenario to explain the tree protection operation initiated by a leaf node. Initially, the network is in a normal condition and traffic flows on the working tree. When a leaf node, for example *L*1, detects an SF on the working path, all the root and leaf nodes will switch to the protection tree. The operational sequence for this example scenario is illustrated in Fig. 6(b) and summarized as follows:

1) All the root and leaf nodes are operating under normal conditions and using the working tree (w) for traffic delivery,

which is indicated by the "no request" NR(w) messages in the figure.

2) *L*1 detects an SF on working.
3) *L*1 switches to the protection tree (p) by setting the bridge and selector (b/s) to the protection tree, and sends an "SF(p)" message to the root node.
4) The root node switches to the protection tree and sends a new message, "X(p)," to all the leaf nodes (*L*1, … , *LN*).
5) Upon receiving message "X(p)" from the root node, *L*1 takes no action and keeps sending "SF(p)," but all the other leaf nodes, *L*2, … , *LN*, switch to the protection tree and send confirmation message "NR(p)" to the root node.

The purpose of the new message "X(p)" is to request protection switching to all the leaf nodes bar the one that actually initiated the protection switching. For the leaf node who initiated the protection switching (*L*1 in this example), the message "X(p)" can be considered as a confirmation. In a tree protection architecture, the protection switching action is determined by the request having the highest priority within the protected domain. For this, the root node should propagate the highest priority request in the protected domain by comparing the priority of remote requests from each leaf node with its local request (top priority local request), and sending the highest one (top priority global request) to each leaf node. In the above example, the message "X(p)" should be "SF(p)," which is deviated from the existing APS protocol. In the APS protocol, the root node would send "NR(p)" as it has no defect detected locally and sends and receives traffic to and from the protection tree.

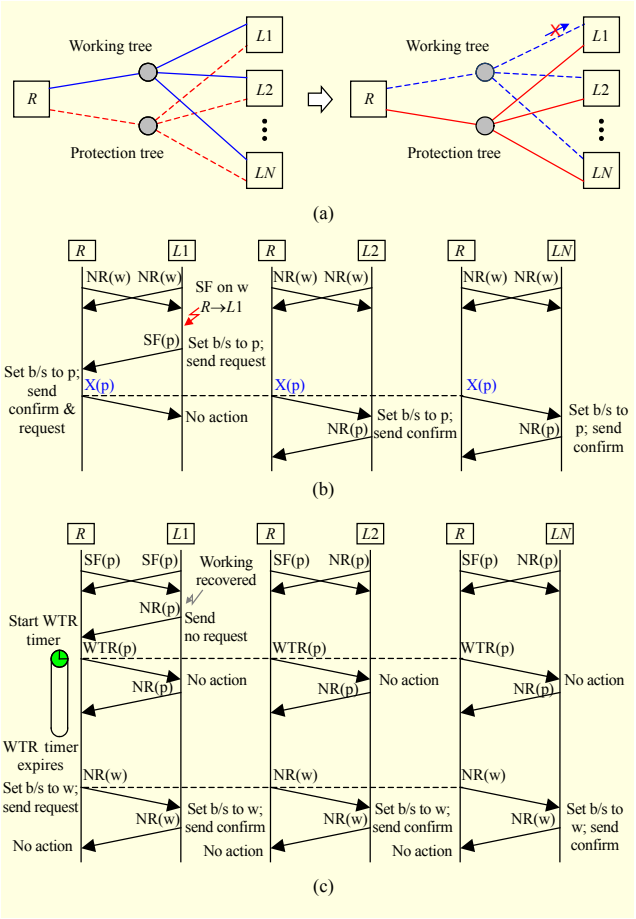

Fig. 6. Tree protection operation initiated by leaf node: (a) architecture, (b) sequence of protection switching operation, and (c) sequence of reversion operation.

According to the current operational principle of APS Ethernet linear protection protocol, a node sends the top priority global request only when it is the top priority local request, and sends "NR(p)" when the top priority global request is a remote request from the far-end node. A root node always sends its top priority global request regardless of its origin. It should be noted that this change suggested in this paper can cause leaf nodes to recognize the propagated request from the root node as the local request of the root node. In the above example, all the leaf nodes will consider the received "SF(p)" as if the root node detected an SF in the direction of any leaf node to the root node. In bi-directional protection switching, this forged SF message at a leaf node does not affect the Ethernet linear protection APS protocol operation in the leaf node except for the recovery case.

Figure 6(c) is a continuation of Fig. 6(b) and shows the operational sequence for the reversion case when the working tree is recovered from the failure. The operational sequence of Fig. 6(c) is summarized as follows:

1) All the root and leaf nodes are in protection mode due to an SF on the working tree at *L*1 (root node is propagating "SF(p)" to all the leaf nodes).
2) *L*1 detects clearance of SF on working tree.
3) *L*1 sends "NR(p)" to the root node.
4) Upon receiving "NR(p)," the root node starts a wait-to-restore (WTR) timer, which is used to avoid chattering of selectors in the case of intermittent defects, and sends "WTR(p)" to all the leaf nodes.
5) When the WTR timer expires, the root node switches to the working tree and sends "NR(w)" to all the leaf nodes.
6) Upon receiving "NR(w)," each leaf node switches to the working tree and sends a confirmation "NR(w)" to the root node.

It should be noted that the root node runs a WTR timer when its top priority global request is changed from an SF to a "no request" despite the fact that the SF is not its local request. A leaf node does not start its WTR timer as it does not receive "NR(p)" from the root node after the recovery

B. Case of Detecting SFs at Root Node

If a root node initiates protection switching, then tree

ETRI Journal, Volume 38, Number 1, February 2016
http://dx.doi.org/10.4218/etrij.16.0115.0024

protection will be completed by a single message exchange between the root node and all the leaf nodes. In this case, the APS Ethernet linear protection protocol can be used without any modifications. Figure 7(a) shows an example scenario to explain the tree protection operation initiated by a root node. Initially, the protected domain is in a normal condition and traffic is flowing on the working tree (Fig. 7(a)). When a root node detects an SF on the working path from a leaf node (for example *L*1) to the root node, all the root and leaf nodes will switch to the protection tree (Fig. 7(b)). The operational sequence for this example scenario is illustrated in Fig. 7(b) and summarized as follows:

1) Initially, all the root and leaf nodes are in normal condition, which is indicated by the presence of "NR(w)" messages.
2) The root node detects an SF on the working tree.
3) The root node switches to the protection tree and sends a request "SF(p)" to all the leaf nodes.
4) Each leaf node switches to the protection tree and sends "NR(p)" to the root node.

When the working tree recovers from the failure, the protected domain will revert to the working tree after WTR timer expiration if the domain is configured with a revertive mode, or keep the protection tree if the domain is configured with a non-revertive mode. As a continuation of Fig. 7(b), Fig. 7(c) shows the operational sequences for the case of revertive mode. The operational sequence of Fig. 7(c) is summarized as follows:

1) All the root and leaf nodes are in protection mode due to an SF on the working tree at the root node.
2) The root node detects clearance of the SF on the working tree.
3) The root node starts its WTR timer and sends "WTR(p)" to all the leaf nodes.
4) When the WTR timer expires, the root node switches to the working tree and sends "NR(w)" to all the leaf nodes.
5) Upon receiving an "NR(w)," each leaf node switches to the working path and sends a confirmation "NR(w)" to the root node.

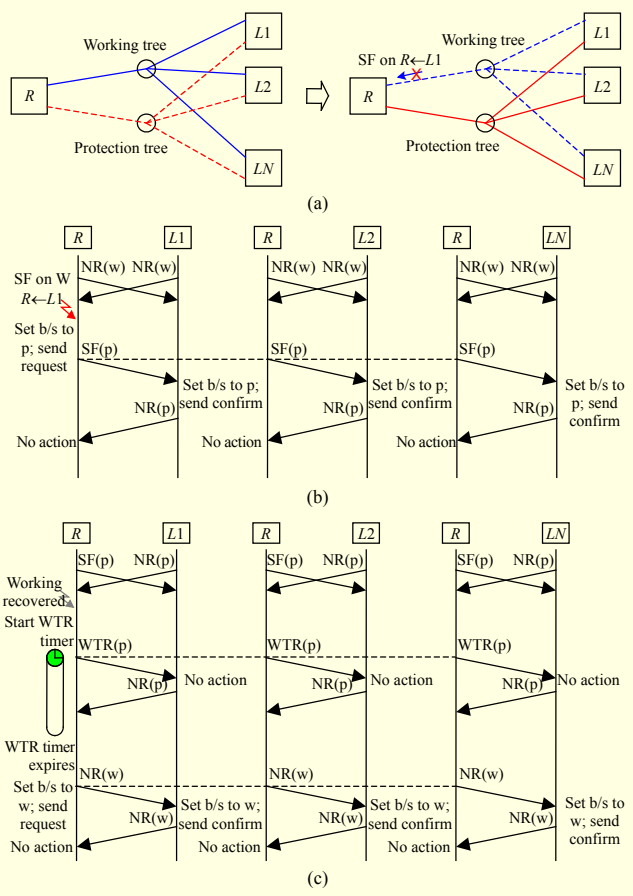

Fig. 7. Tree protection operation initiated by root node: (a) architecture, (b) sequence of protection switching operation, and (c) sequence of reversion operation.

4. Hybrid Scheme of Per-leaf and Tree Protections

This protection mechanism allows flexible operational changes between per-leaf and tree protection schemes. When a failure happens near leaf nodes or some quantity of multiple failures occur, a per-leaf protection mechanism is used. If a defect occurs on a link near the root node or many connections on the working tree are affected at the same time, then it may not be able to achieve protection switching within 50 ms due to the significant amount of APS processing burden. At a time when the number of leaf nodes being affected by a defect or the number of protection processes that are running simultaneously exceeds a certain threshold within a certain period of time, per-leaf protection is suspended and tree protection is initiated to reduce APS processing burden. However, when a prior failure exists in one tree, and the second failure occurs in the other tree, per-leaf protection is activated to enhance the network availability by utilizing resources on both trees.

III. Performance Evaluation

In this section, we introduce a real test bed in which the proposed schemes are implemented and measure their performances. Then, based on our experience with the real system, we formulate the restoration times for various schemes and compare their performances. An OPNET simulator [15] is used to evaluate the performance of tree protection.

1. Restoration Time of Per-leaf Protection

Resilience is a key attribute of PTNs, and the transfer time must satisfy a sub–50 ms SONET/SDH-grade resiliency [16].

Protected traffic restoration time (T_P) is the time from the occurrence of the network impairment to the restoration of protected traffic and can be expressed as

$$T_P = T_C + T_T + T_R, \quad (1)$$

where T_C (confirmation time) is the time from the occurrence of the network impairment to the instant when the triggered SF is confirmed as requiring protection switching operations. It is the sum of the detection and hold-off times. The transfer time, T_T, is the time interval between the confirmation of the SF and the completion of the protection switching operations, which include setting up the positions of bridge and selector and transmitting any resulting protection protocol message. The recovery time, T_R, is the time interval between the completion of protection switching operations and the full restoration of protected traffic; this includes the verification of switching operations [17]. Although the sub–50 ms protection switching time requirement is normally applied to T_T in standards [1], [6], our analysis and experiments focus on T_P to evaluate the performances of various schemes from the aspect of service traffic recovery as a whole.

To demonstrate the survivability performances of per-leaf protection using Ethernet linear protection with or without C-SF, a real test-bed network is built as in Fig. 8. The experimental network consists of one root node and two aggregation leaf nodes that run our proposed schemes, and two intermediate nodes of commercial Ethernet switches. The root and leaf nodes are developed based on Broadcom Petra-B, FE600, Altera Arria II GX, and MPC8543V CPU with Linux 2.6.35. One 10GBASE-LR Ethernet link is used between a root node and an intermediate node and sixty 1000BASE-SX links are placed between each aggregation leaf node and one of two intermediate nodes. Each leaf node has 10 line-cards — five of which belong to the working tree and the rest to the protection tree. Each line-card is connected to intermediate nodes with 12 physical links. The root node is configured to have up to 1,200 LPPs. Each aggregation leaf node can handle up to 600 LPPs. The aggregation leaf node in this experimental network serves 600 clients, each of which acts as the leaf node in Fig. 1.

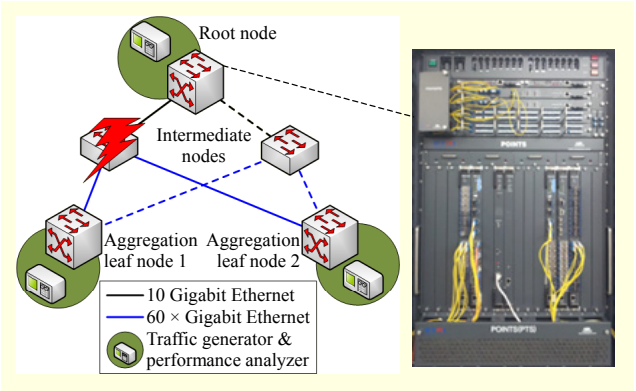

Fig. 8. Experimental setup.

For user traffic for each LPP, the traffic generator generates Ethernet frames at an average rate of 10 kfps with 640 bits per frame. All the user traffic is assumed to be symmetric and co-routed so that the same amounts of traffic between source and destination are delivered through the same set of links and nodes in each direction. The length of each link is less than 10 m.

In our system, a processor is assigned to one task at any time and the task is executed for one time slice, whose length is a given number of the jobs that the task can process at one time slice. All the LPPs in a node are handled by one task, which is called LPP task. The length of one time slice varies in our experiments. We consider two kinds of queues: the input queue for the LPP task where the received SF notifications are placed in and the output queue for the MEP task that detects LOC events and sends SF notifications. In our multi-slot system, two tasks reside in separate line cards. The number of SF notifications processed by the LPP task in one time slice is called Q_{in}, whereas the number of SF notifications that the MEP task sends via IPC in one time slot is called Q_{out}. The time interval between the completion of one time slice and the start of the next time slice for the LPP task and the MEP task, T_{Qin} and T_{Qout}, respectively, are measured to be 1 ms, on average.

In our implementation, as the reception of remote SF messages happens at the same line-card as the LPP task resides, no IPC is assumed for the SF messages sent from the remote end node.

As shown in Fig. 9, the restoration time increases with the number of LPPs in a root node. It also demonstrates that the increase of Q_{in} improves the performance of protection

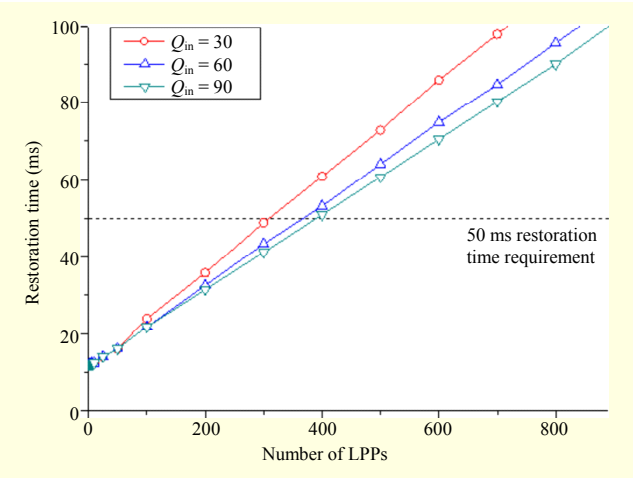

Fig. 9. Restoration time comparison for three durations of time slice for LPP task.

ETRI Journal, Volume 38, Number 1, February 2016
http://dx.doi.org/10.4218/etrij.16.0115.0024

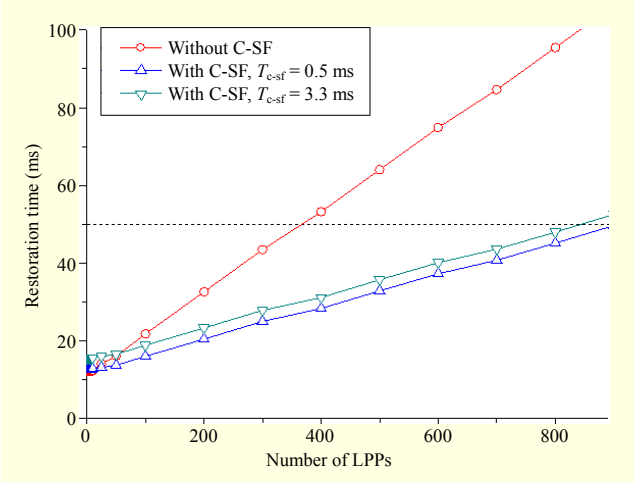

Fig. 10. Restoration time for C-SF mechanism.

switching. The number of LPPs that meet the 50 ms restoration time requirement increases from 300 to 400 when Q_{in} increases from 30 to 90. The value of Q_{out} is fixed to 25 in this experiment.

The performances of per-leaf protection with and without the proposed C-SF mechanism are shown in Fig. 10. In the C-SF mechanism, all the SFs that occur during a certain time interval are collected in one notification and the time interval is called $T_{c\text{-}sf}$. When the number of SFs exceeds the maximum number of SFs conveyed in one notification, $M_{c\text{-}sf}$, a notification is generated without waiting for $T_{c\text{-}sf}$. For the experimental results shown in Fig. 10, $T_{c\text{-}sf}$ is set to either 0.5 ms or 3.3 ms, and $M_{c\text{-}sf}$ is set to 200. The value of Q_{in} is set to 60. The result demonstrates that the C-SF mechanism reduces the restoration time significantly. The number of LPPs that meet the 50 ms restoration time requirement increases from 380 to 900 when the proposed C-SF mechanism is used with 0.5 ms of $T_{c\text{-}sf}$.

2. Formulation of Restoration Time for Per-leaf Protection with Linear Protection Algorithms

Based on our experience with the system in which the proposed schemes are implemented, we formulate the restoration time with simultaneous multiple protection triggers for per-leaf protection with linear protection algorithms assuming the network impairment occurs at the working tree. In addition to the notations introduced in previous sections, we use the following notations:

- N: Total number of LPPs at the root node. The same number as the number of leaf nodes, each of which has one LPP.
- n: Number of the affected LPPs that require traffic switchover simultaneously, $1 \le n \le N$.
- $T_P(i, n)$: Protected traffic restoration time of the ith LPP among n affected LPPs. The value is the same for the peering leaf node to the ith LPP.
- $T_{DR}(i, n)$: Time between the occurrence of network impairment and the detection of an SF at the ith working MEP among n affected working MEPs in the root node. When the latency of a CC message [14] from the impairment location to the root node is T_{prop_r} and the CC message interval is CC_Period, the value of $T_{DR}(i, n)$ is in between $2.5 \times$ CC_Period + T_{prop_r} and $3.5 \times$ CC_Period + T_{prop_r}.
- $T_{DL}(i)$: Time between the occurrence of network impairment and the detection of an SF at the working MEP in leaf node i, which corresponds to the ith working MEP in the root node. When the latency of a CC message [14] from the impairment location to the leaf node i is T_{prop_i} and the CC message interval is CC_Period, the value of $T_{DL}(i)$ is in between $2.5 \times$ CC_Period + T_{prop_i} and $3.5 \times$ CC_Period + T_{prop_i}.
- $T_H(i, n)$: Hold-off time interval of the ith LPP among n affected LPPs. As the same hold-off time value is used in both end nodes of a P2P connection, an LPP of a leaf node has the same hold-off time interval as its peer LPP in the root node.
- $T_{ipcR}(i, n)$: IPC processing time of the local SF detected at the ith working MEP among n affected MEPs in the root node.
- $T_{ipcL}(i)$: IPC processing time of the local SF detected at the working MEP in leaf node i, which corresponds to the ith working MEP in the root node.
- $T_{TR}(i, n)$: Transfer time triggered by either a local SF or remote SF message at the ith LPP of the root node. Both local SFs and remote SF messages are processed in the order they arrive. This time is consumed by the root node LPP task that performs the protection switching operation for the SF.
- $T_{TL}(i)$: Transfer time triggered by either a local SF or remote SF message at the LPP of leaf node i, which corresponds to the ith LPP of the root node. This time is consumed by the leaf node LPP task that performs the protection switching operation for the SF.
- $T_{R2L}(i, n)$: End-to-end delay of a protection protocol message from the ith LPP of the root node to the LPP of leaf node i, which corresponds to the ith LPP of the root node. This value is related to the transmission speeds and length of links and the packet processing times in intermediate nodes.
- $T_{L2R}(i, n)$: End-to-end delay of a protection protocol message from the LPP of leaf node i, which corresponds to the ith LPP of the root node. This value is related to the transmission speeds and length of links and the packet processing times in intermediate nodes.
- $\lfloor\ \rfloor$ indicates the floor function.

Then, protected traffic restoration time, T_P, is expressed as

$$T_P = \max_i T_P(i, n). \quad (2)$$

Depending on the types of SFs, $T_P(i, n)$ can be calculated as in the following subsections.

A. Case of Unidirectional SFs Detected at Root Node

When multiple unidirectional SFs occur in the direction from the leaf nodes to the root node,

$$T_{\mathrm{P}}(i,n) = T_{\mathrm{DR}}(i,n) + T_{\mathrm{H}}(i,n) + \sum_{x=i-a_i}^{i} T_{\mathrm{ipcR}}(x,n) + \left\lfloor \frac{a_i}{Q_{\mathrm{out}}} \right\rfloor \times T_{\mathrm{Qout}} + \sum_{x=i-b_i}^{i} T_{\mathrm{TR}}(x,n) + \left\lfloor \frac{b_i}{Q_{\mathrm{in}}} \right\rfloor \times T_{\mathrm{Qin}} + T_{\mathrm{R2L}}(i,n) + T_{\mathrm{TL}}(i) + T_{\mathrm{L2R}}(i,n), \quad (3)$$

where a_i is the number of SF notifications to be sent by the MEP task when the ith SF is detected; b_i is the number of SF notifications waiting to be processed by the LPP task at the time that the ith SF notification arrives at the input queue of the LPP task.

B. Case of Unidirectional SFs Detected at Leaf Nodes

When multiple unidirectional SFs occur in the direction from the root node to the leaf nodes,

$$T_{\mathrm{P}}(i,n) = T_{\mathrm{DL}}(i) + T_{\mathrm{H}}(i,n) + T_{\mathrm{ipcL}}(i) + T_{\mathrm{TL}}(i) + T_{\mathrm{L2R}}(i,n) + \sum_{x=i-c_i}^{i} T_{\mathrm{TR}}(x,n) + \left\lfloor \frac{c_i}{Q_{\mathrm{in}}} \right\rfloor \times T_{\mathrm{Qin}} + T_{\mathrm{R2L}}(i,n), \quad (4)$$

where c_i is the number of SFs waiting to be processed by the LPP task of the root node at the time that the SF message from leaf node i arrives at the input queue of the LPP task.

C. Case of Bidirectional SFs Detected at Both Root Node and Leaf Nodes

When multiple bidirectional SFs occur in both directions,

$$T_{\mathrm{P}}(i,n) = T_{\mathrm{H}}(i,n) + \max\left\{ \left(T_{\mathrm{DR}}(i,n) + \sum_{x=i-a_i}^{i} T_{\mathrm{ipcR}}(x,n) + \left\lfloor \frac{a_i}{Q_{\mathrm{out}}} \right\rfloor \times T_{\mathrm{Qout}} + \sum_{x=i-c_i}^{i} T_{\mathrm{TR}}(x,n) + \left\lfloor \frac{c_i}{Q_{\mathrm{in}}} \right\rfloor \times T_{\mathrm{Qin}} + T_{\mathrm{R2L}}(i,n) \right), \left(T_{\mathrm{DL}}(i) + T_{\mathrm{ipcL}}(i) + T_{\mathrm{TL}}(i) + T_{\mathrm{L2R}}(i,n) \right) \right\}. \quad (5)$$

3. Restoration Time Analysis for Per-leaf Protection with C-SF Mechanism

In this section, we consider the restoration time of per-leaf protection with the proposed C-SF mechanism. Assuming the SF detected at the ith working MEP among n affected working MEPs in the root node is collected in the jth collective SF notification, $T_{\mathrm{P}}(i,n)$ for per-leaf protection with C-SF mechanism when multiple bidirectional SFs occur in both directions, is expressed as (6). In (6), $T_{\mathrm{Wc\text{-}sf}}(i, n)$ is the time between the detection of an SF at the ith working MEP among n affected working MEPs in the root node and the completion of the jth collective SF notification ready to be sent out via IPC, $0 \le T_{\mathrm{Wc\text{-}sf}}(i, n) \le T_{\mathrm{c\text{-}sf}}$. The number of collective SF notifications to be sent by the MEP task when the jth collective SF notification is formed is represented by d_i.

$$T_{\mathrm{P}}(i,n) = T_{\mathrm{H}}(i,n) + \max\left\{ \left(T_{\mathrm{DR}}(i,n) + T_{\mathrm{Wc\text{-}sf}}(i,n) + \sum_{x=j-d_j}^{j} T_{\mathrm{ipcR}}(x,n) + \left\lfloor \frac{d_j}{Q_{\mathrm{out}}} \right\rfloor \times T_{\mathrm{Qout}} + \sum_{x=i-c_i}^{i} T_{\mathrm{TR}}(x,n) + \left\lfloor \frac{c_i}{Q_{\mathrm{in}}} \right\rfloor \times T_{\mathrm{Qin}} + T_{\mathrm{R2L}}(i,n) \right), \left(T_{\mathrm{DL}}(i) + T_{\mathrm{ipcL}}(i) + T_{\mathrm{TL}}(i) + T_{\mathrm{L2R}}(i,n) \right) \right\}. \quad (6)$$

We omit the expressions of $T_{\mathrm{P}}(i,n)$ for the cases of unidirectional SFs, as they can easily be derived similarly.

Figure 11 shows the restoration times calculated from (5) and (6) with various values of Q_{in} and $T_{\mathrm{c\text{-}sf}}$. We can observe that the graphs match the experimental results shown in Figs. 9 and 10. We assume that $T_{\mathrm{H}}(i, n)$ is zero and $T_{\mathrm{DR}}(i, n)$ occurs equally likely between 2.5 × CC_Period + $T_{\mathrm{prop\_r}}$ and 3.5 × CC_Period + $T_{\mathrm{prop\_r}}$. From our experiments, the mean of $T_{\mathrm{ipcR}}(i, n)$ shows two different values for $i = 1$ and $i \ge 2$ and the values measure 29 μs and 19 μs for $i = 1$ and $i \ge 2$, respectively. The mean of $T_{\mathrm{TR}}(i, n)$, measure 87 μs, 32 μs, or 21 μs when i(mod Q_{in}) is 1, 2, or ≥ 3, respectively.

4. Restoration Time of Tree Protection and its Comparison with other Protection Protocols

For tree protection, the protected traffic restoration time for

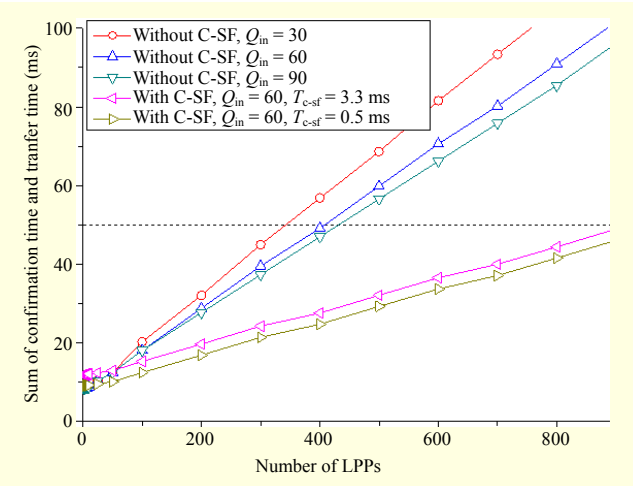

Fig. 11. Restoration time comparison based on derived equations.

ETRI Journal, Volume 38, Number 1, February 2016
http://dx.doi.org/10.4218/etrij.16.0115.0024

unidirectional SFs detected at the root node can be written as

$$T_{\mathrm{P}} = T_{\mathrm{DR}}(1,n) + T_{\mathrm{H}} + T_{\mathrm{ipcR}}(1,n) + T_{\mathrm{TR}} + \max_{i}\left\{T_{\mathrm{R2L}}(i,n) + T_{\mathrm{TL}}(i) + T_{\mathrm{L2R}}(i,n)\right\}, \quad (7)$$

where T_{H} and T_{TR} denote the hold-off time interval and the transfer time at the TPP in the root node, respectively.

The restoration time of tree protection for unidirectional SFs detected at the leaf nodes is written as

$$T_{\mathrm{P}} = T_{\mathrm{H}} + \min_{i}\left\{T_{\mathrm{DL}}(i) + T_{\mathrm{ipcL}}(i) + T_{\mathrm{TL}}(i) + T_{\mathrm{L2R}}(i,n)\right\} + T_{\mathrm{TR}} + \max_{j}\left\{T_{\mathrm{R2L}}(j,n) + T_{\mathrm{TL}}(j) + T_{\mathrm{L2R}}(j,n)\right\}. \quad (8)$$

We leave formulation of the restoration time of tree protection for bidirectional SFs for future work.

The performance of protection switching with a tree protection protocol and its benefits are evaluated in comparison with other protection switching protocols in P2MP connection networks. To examine transient behaviors of different protection protocols more closely, we rely on an OPNET simulator. As shown in Fig. 12, our computer simulation network consists of one root node (*R*) with a server (*S*), 42 intermediate nodes, and 1,000 leaf nodes, each of which is connected to a client host. All the links connecting the server, the root node, and intermediate nodes are assumed to have a 10 Gbps capacity and the remaining links are assumed to have a 1 Gbps capacity, which is large enough not to limit the traffic volume generated by the server and each client host. Any pairs of two adjacent nodes are connected with 80 km fiber links, whose propagation delay is approximated to 400 μs.

Each host generates Ethernet frames at an average rate of 5 kfps destined to the server. The server generates Ethernet frames at an average rate of 5 kfps for each client host. The lengths of Ethernet frames are exponentially distributed with a mean of 100 bytes and their inter-arrival times are also exponentially distributed. The CC message rate is always set to 333 fps. According to the standard protocol message transmission rule [1]–[6], APS or R-APS message transmission rates are set to 333 fps in a burst mode, which is supposed to happen right after the protection switching occurs, followed by a continuous mode, where the messages are generated at every 5 s. In the case of hybrid protection, the threshold value for from per-leaf to tree protection is set to 350 protection processes within a 6.6 ms interval.

The two graphs in Fig. 13 show the restoration times for all the aforementioned protection schemes in the cases of bidirectional SFs and unidirectional SFs in a P2MP connection network. The results are obtained by varying the number of leaf nodes that are affected simultaneously by a single network defect. For ring protection, all the leaf nodes are assumed to be RPL owner, and the RPL blocks are placed at the links in the working tree. The restoration time of ring protection is longer

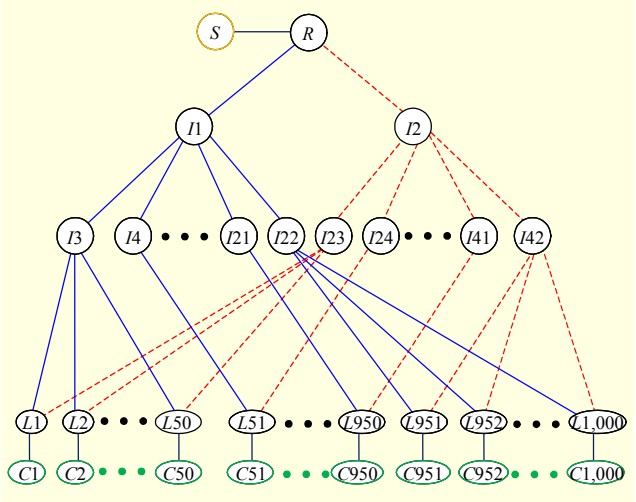

Fig. 12. Simulation scenario in normal Ethernet tree topology.

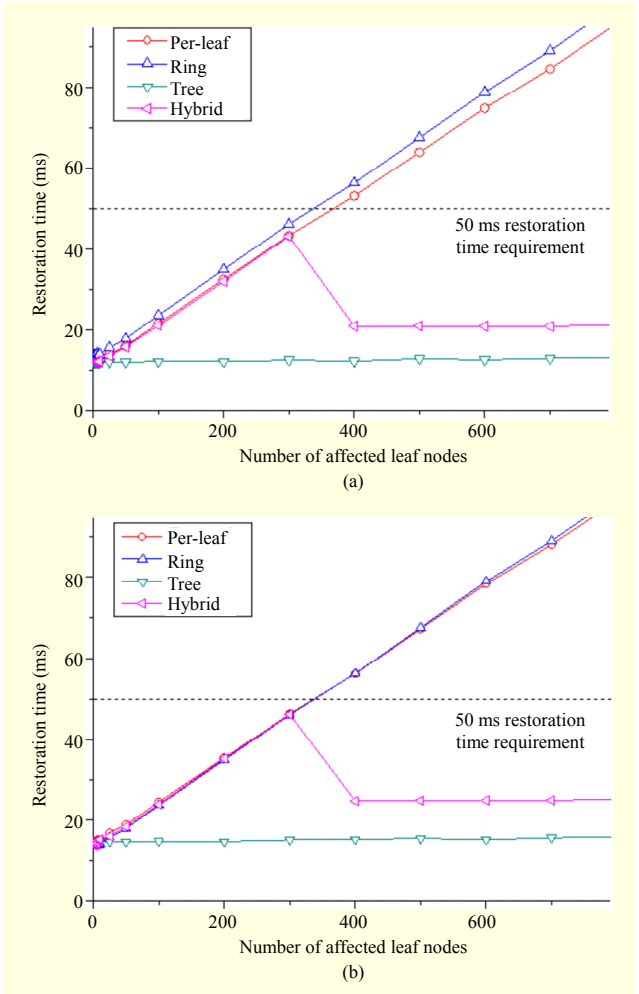

Fig. 13. Restoration time comparison: (a) bidirectional SF and (b) unidirectional SF.

than that of per-leaf protection because of the additional delay of R-APS. Both ring and per-leaf protection schemes show that the restoration times increase as the number of affected leaf nodes increases. In contrast, tree and hybrid protection schemes are stabilized regardless of the number of affected leaf nodes. The performance of the hybrid scheme in terms of restoration time cannot be better than that of tree protection, but it can utilize resources on both trees and enhance the network availability more so than tree protection if the affected leaf nodes are less than the threshold value.

When all the 1,000 leaf nodes are affected by a link failure, the data rates measured at the link from the root node to the server and the links from the leaf nodes to their client hosts are shown in Fig. 14. The link failure occurs at 1.0 s, and the data rates are measured every 5 ms. Traffic is recovered in about 20 ms and 30 ms after the failure for the tree and hybrid protection schemes, respectively. However, in the cases of ring and per-leaf protections, their results exceeded 100 ms.

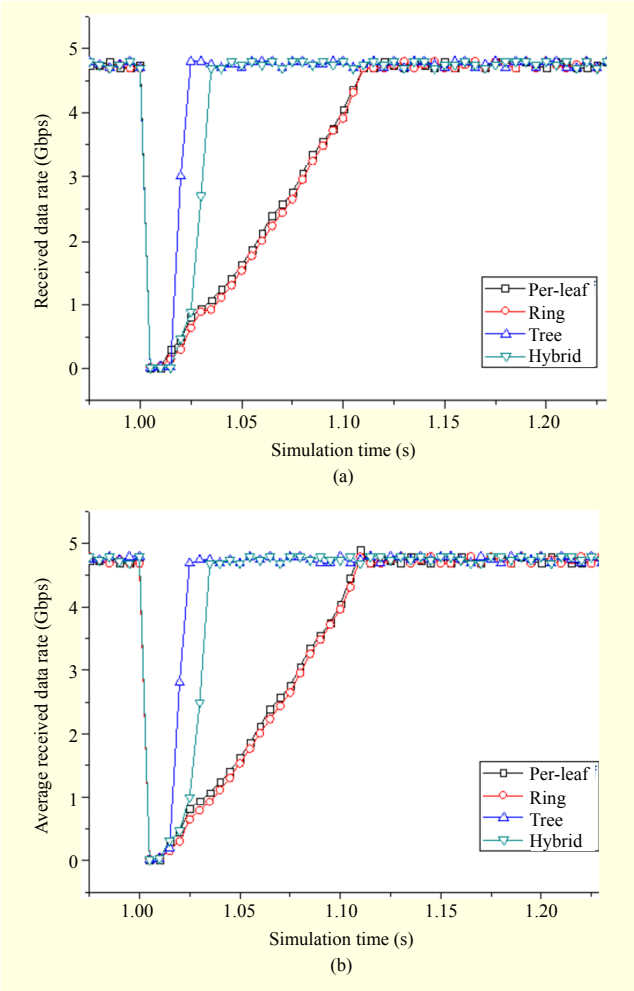

Fig. 14. Single link failure in P2MP topology: (a) data rate at link from *R* to *S* and (b) average data rate at each link from leaf node to client.

IV. Conclusion

We proposed various protection switching schemes for P2MP connections in PTN. The main purpose of a P2MP connection protection is to achieve fast traffic recovery while the existing protection switching technologies can be reused with minimal modifications.

Observing that the IPC time has a strong influence on the restoration time, a C-SF mechanism was proposed to relieve the burden related to IPC for SF notifications in the case of a per-leaf protection scheme. To maximize the agility of protection switching for the P2MP connection, tree protection was considered and its detailed operational behavior was presented. To maximize the network availability subject to the sub–50 ms protection switching time constraint, a hybrid of per-leaf and tree protection schemes was also proposed.

A comparison study among the presented schemes has been performed, and the hybrid scheme showed well-balanced performance between fault-recovery time and network availability. Although the performance of per-leaf protection schemes largely depends on the number of protected instances in a P2MP connection, the tree and hybrid schemes guarantee fast and reliable protection switching regardless of the number of leaf nodes in a P2MP connection network.

References

[1] ITU-T Rec. G.8031/Y.1342, *Ethernet Linear Protection Switching*, June 2011.

[2] IETF RFC 7271, *MPLS Transport Profile (MPLS-TP) Linear Protection to Match the Operational Expectations of Synchronous Digital Hierarchy, Opt. Transport Network, and Ethernet Transport Network Operators*, June 2014.

[3] ITU-T Rec. G.8131/Y.1382, *Linear Protection Switching for MPLS Transport Profile (MPLS-TP)*, July 2014.

[4] IETF RFC 6378, *MPLS Transport Profile (MPLS-TP) Linear Protection*, Oct. 2011.

[5] J.-D. Ryoo et al., "MPLS-TP Linear Protection for ITU-T and IETF," *IEEE Commun. Mag.*, vol. 52, no. 12, 2014, pp. 16–21.

[6] ITU-T Rec. G.8032/Y.1343, *Ethernet Ring Protection Switching*, Feb. 2012.

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

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

[9] K.-K. Lee, J.-D. 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.

[10] K.-K. Lee, J.-D. Ryoo, and Y. Kim, "Impacts of Hierarchy in Ethernet Ring Networks on Service Resiliency," *ETRI J.*, vol. 34, no. 2, Apr. 2012, pp. 199–209.
[11] K.-K. Lee and J.-D. Ryoo, "Flush Optimization to Guarantee Less Transient Traffic in Ethernet Ring Protection," *ETRI J.*, vol. 32, no. 2, Apr. 2010, pp. 184–194.
[12] K.-K. Lee, C.-K. Lee, and J.-D. Ryoo, "Enhanced Protection Schemes to Guarantee Consistent Filtering Database in Ethernet Rings," *IEEE Global Commun. Conf.*, Miami, FL, USA, Dec. 6–10, 2010, pp. 1–6.
[13] The Metro Ethernet Forum MEF 10.2, *Ethernet Services Attributes Phase 2*, Oct. 2009.
[14] J.-D. 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. Accessed Oct. 2012. http://www.opnet.com
[16] M. Huynh, S. Goose, and P. Mohapatra, "Resilience Technologies in Ethernet," *Comput. Netw.*, vol. 54, no. 1, Jan. 2010, pp. 57–78.
[17] ITU-T Rec. G.808.1, *Generic Protection Switching – Linear Trail and Subnetwork Protection*, May 2014.

Dae-Ub Kim is a principal researcher at ETRI and is currently working toward a PhD degree in information and communications engineering at Chungnam National University, Daejeon, Rep. of Korea. He received his MS degree in information and communications engineering from the Korea Advanced Institute of Science and Technology, Daejeon, Rep. of Korea, in 2001 and his BS degree in electronic engineering from Yeungnam University, Gyeongsan, Rep. of Korea, in 1999. Since he joined ETRI in 2001, his work has been focused on next-generation networks, wireless backhaul networks, carrier-class Ethernet, OTN, and MPLS-TP technology research, especially participating in protection standardization activities in ITU-T.

Jeong-dong Ryoo is a principal researcher at ETRI and a UST professor with the Department of Engineering, Korea University of Science and Technology, Daejeon, Rep. of Korea. He holds MS and PhD degrees in electrical engineering from the Polytechnic Institute of New York University, USA and a BS degree in electrical engineering from Kyungpook National University, Daegu, Rep. of Korea. Upon completing his PhD in the area of telecommunication networks and optimization, he started working for Bell Labs, Lucent Technologies, 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-class Ethernet, and MPLS-TP technology research, especially participating in OAM and protection standardization activities in ITU-T. He is the editor of G.8131 (MPLS-TP linear protection), G.8132 (MPLS-TP ring protection), and G.808.1 (Generic protection - Linear) recommendations and a vice-chairman of ITU-T Study Group 15. He co-authored *TCP/IP Essentials: A Lab-Based Approach* (Cambridge University Press, 2004). He is a member of the Eta Kappa Nu association.

Jong Hyun Lee received his BS, MS, and PhD degrees in electronics engineering from Sungkyunkwan University, Suwon, Rep of Korea, in 1981, 1983, and 1993, respectively. Since 1983, he has been with ETRI, where he has served as a director for both the Optical Communication Department and the Research Strategy & Planning Department. He has also served as an executive director of the Optical Internet Research Department. His current research interests are packet-circuit-optical converged switching systems, green Internet data centers, and optical access networks.

Byung Chul Kim received his BS degree in electronic engineering from Seoul National University, Rep. of Korea and his MS and PhD degrees in electronic engineering from the Korea Advanced Institute of Science and Technology, Daejeon, Rep. of Korea, in 1988, 1990, and 1996, respectively. From 1993 to 1999, he worked as a research engineer at Samsung Electronics, Suwon, Rep. of Korea. Since 1999, he has been a professor at the Department of Information and Communications Engineering, Chungnam National University, Daejeon, Rep. of Korea. His research interests include computer networks, wireless Internet, sensor networks, and mobile communications.

Jae Yong Lee received his BS degree in electronics engineering from Seoul National University, Rep. of Korea and his MS and PhD degrees in electronic engineering from the Korea Advanced Institute of Science and Technology, Daejeon, Rep. of Korea, in 1988, 1990, and 1995, respectively. From 1990 to 1995, he worked as a research engineer at the Digicom Institute of Information and Communications, Seoul, Rep. of Korea. Since 1995, he has been a professor at the Department of Information and Communication Engineering, Chungnam National University, Daejeon, Rep. of Korea. His research interests include computer networks, wireless Internet, sensor networks, and mobile communications.