International Scholarly Research Network
ISRN Communications and Networking
Volume 2011, Article ID 941605, 8 pages
doi:10.5402/2011/941605

Research Article

Fast Fault Recovery Method with Preestablished Recovery Table for Reliable Wide Area Layer-2 Network

Takashi Kurimoto, Midori Terasawa, Sho Shimizu, Satoru Okamoto, and Naoaki Yamanaka

Department of Information and Computer Science, Faculty of Science and Technology, Keio University, 3-14-1 Hiyoshi, Kohoku-ku, Yokohama 223-8522, Japan

Correspondence should be addressed to Satoru Okamoto, okamoto@yamanaka.ics.keio.ac.jp

Received 28 June 2011; Accepted 1 August 2011

Academic Editor: M. Listanti

Recently, wide area Ethernet, which provides virtual connections between company branches by using Ethernet technology, has become wide spread. Ethernet originally lacked fault recovery mechanisms required in wide area networks (WANs) since it originated as local area network (LAN) technology. Protection mechanism against link or node failure is proposed in Ethernet network, but many frame losses are inevitable during the switch over to the backup path. Therefore, to improve reliability of wide area Ethernet, a fault recovery method with few frame losses is required. This paper proposes a fast fault recovery method for reducing frame losses during the switching over to another route. To achieve fast recovery, a pre-established forwarding table is introduced, which is searched when a fault occurs and frames are forwarded. This paper also proposes a bandwidth control function to guarantee the quality of service of connections that do not use the failed link when a fault occurs. A computer simulation showed that by adding the proposed fault recovery method to the conventional protection method, frame losses can be decreased. The proposed fault recovery method was implemented on a software switch framework to confirm its effectiveness.

1. Introduction

Recently, Ethernet virtual private network (VPN) services provided by carriers have been gaining attention regarding high-speed communication among local area network (LAN) sites with Gigabit Ethernet (GbE) and 10-Gigabit Ethernet (10 GbE) [1].

Ethernet VPN services enable service providers to provide site-to-site Ethernet connectivity (c.f. MEF E-line) [2] and multisite Ethernet connectivity (c.f. MEF E-LAN). Unlike IP VPN services, using Ethernet as the service delivery technology often results in lower operation expense (OpEx) when compared to VPN services delivered using IP or IP/MPLS technologies.

Users can extend a LAN to another site simply through an Ethernet VPN without buying any additional routers. Moreover, the standardization of 40 Gb/100 Gb Ethernet is also advanced [1] and expected to connect wide area networks (WANs) in the future.

However, the Ethernet switches for an Ethernet VPN were originally designed for a LAN. Thus, there is a lack of functions when Ethernet switches are applied to a WAN. Such WAN's overall performance is not as good as that of legacy-leased line services. However, Ethernet technologies have been improved for application to WANs, and many functions, such as fault recovery, quality of service (QoS) [3, 4], and OAM, have been proposed and developed for wide area Ethernet.

We focus on fault recovery for reliable Ethernet VPN. Cloud computing technology and various applications using cloud computing technology, including mission critical applications, have been developed and will work over an Ethernet VPN. Thus, high reliability of an Ethernet VPN has become a more important requirement.

A fast fault recovery function is an important technology for high availability. A fast recovery mechanism works in a WAN with IP/MPLS technology. For example, with local repair with a pre-established path, the downtime after failure will be shorter than a few tens of milliseconds.

Fast recovery mechanisms of Ethernet, such as resilient packet ring (RPR) and Ethernet ring protection (ERP), have

been developed, but these mechanisms can be applied only for ring topology, not for mesh topology. Therefore, fast recovery mechanisms for a mesh topology Ethernet network are required.

We, therefore, propose a new fast recovery method for E-line service in an Ethernet VPN. This mechanism is based on local repair with a pre-established path. With this method, downtime after failure will be shortened to a few milliseconds.

We also introduce a new pre-established forwarding table (hereafter called "recovery table") used during failure.

Each node has its recovery table and each recovery table is established by Generalized Multi-Protocol Label Switching (GMPLS) [5] control plane. This table is renewed when a network topology is changed.

During normal operation, Ethernet switches use the virtual LAN ID (VLAN-ID) of Ethernet frames and select the forwarding port. Each forwarding table from ingress node to egress node is set up, and a virtual circuit is established in the Ethernet network, the same as MPLS, asynchronous transfer mode (ATM), and so forth.

When a fault occurs, the Ethernet switch searches for the media access control (MAC) address of the Ethernet frames and determines the forwarding port by searching its recovery table. In other words, frame forwarding continues as soon as a fault is detected.

A computer simulation showed that by adding the proposed fault recovery method to the conventional protection, method, frame losses can be decreased. Additionally, experiments showed that the proposed method achieves fast fault recovery.

The rest of this paper is organized as follows. Section 2 discusses two types of conventional fault recovery methods, protection, and restoration. Section 3 describes our fast fault recovery method with a pre-established recovery table. Section 4 includes the computer simulation results and shows that the proposed method can decrease frame losses when a fault occurs. Section 5 describes the implementation of the proposed method and experimental results. Finally, Section 6 concludes this research.

2. Conventional Fault Recovery Methods

Currently, MPLS technology is often deployed for IP-based WANs. One of the advantages of MPLS is fast fault recovery. It takes a few tens of milliseconds to recover from failure with a fast MPLS recovery mechanism [6]. On the other hand, it normally takes a few minutes for converging the IP route recalculation without MPLS because the routers exchange the fault information and recalculate an alternative route.

There are two repair types: local and global. The node that switches the forwarding path is different between the two types. In local repair, when a fault occurs, the node that detected the fault switches from the working path to the backup path. In global repair, when a fault occurs, the node that detected the fault notifies the ingress node, which switches from the working path to the backup path, as shown in Figure 1.

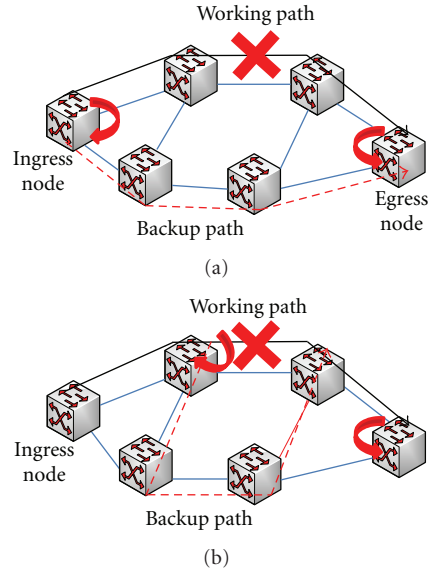

Figure 1: Path recovery in global repair and local repair.

The switching-over time in global repair is slower than that in local repair. In global repair, there are more transactions than in the local repair, for example, generation of notification frames, forwarding of notification frames, and receiving notification frames. However, in local repair, a disadvantage is that the all nodes must have a switching-over function.

There are two types of backup-path route-calculation timings, recovery, and protection.

In protection [7, 8], a backup path is statically reserved for each connection during path set up, and it is used when a fault occurs. In restoration [9], a backup path is dynamically discovered by the node, and the node establishes the backup path and switches from working path to the backup path.

The switching time in restoration is slower than that in recovery because it takes a large amount of time to back up the path route calculation and path configuration after failure. However, network resources are consumed, and multiple faults cannot be recovered in protection.

There are four combinations of the repair types and calculation timings: (1) global repair with protection, (2) global repair with restoration, (3) local repair with protection, and (4) local repair with restoration. Local repair with protection is most preferable for fast recovery.

In an Ethernet network, a forwarding table is generally recovered by the Spanning Tree Protocol/Rapid Spanning Tree Protocol (STP/RSTP) and the MAC address learning mechanisms after failure. By these mechanisms, all nodes have to detect failure and relearn the alternative route after failure; therefore, it takes a few minutes for convergence.

A fast recovery mechanism, traffic engineering provider backbone bridge (PBB-TE), in an Ethernet network has been developed. This mechanism introduces the virtual connection in an Ethernet network. It is also proposed that the mechanism based on global repair with protection in PBB-TE. However, there is no mechanism based on local

repair with protection. Therefore, we propose a fast recovery mechanism based on local repair with protection in an Ethernet network.

3. Proposed Method

This section describes our proposed fast fault recovery method based on local repair with protection for the point-to-point path in E-line service Ethernet networks.

Each node has a newly introduced pre-established recovery table. When a fault occurs, nodes that detect the fault can send frames to the neighboring node by using its recovery table until switching over is completed by global repair. This method reduces the frame losses during switching over by global repair.

3.1. Virtual Connection in Wide Area Ethernet Network. Wide area Ethernet provides virtual connections such as E-line service in an Ethernet VPN. PBB-TE is one of the methods for establishing a virtual connection. At the ingress node of a PBB-TE network, a new Ethernet header is added to the frames.

A VLAN-ID on the newly added header is assigned to each VLAN path, and frames are forwarded with the VLAN-ID. Each node has forwarding tables coupled with each input port, VLAN-ID, and output port. Frames are forwarded to the output port corresponding to the VLAN-ID.

We assumed that wide area Ethernet is controlled by GMPLS for our proposed method. IETF proposed the standardization of the protocol for achieving GMPLS-controlled Ethernet label switching (GELS), which controls the wide area Ethernet by GMPLS. GMPLS is composed of many network control protocols and automates the control to various networks.

In the control plane, a routing protocol, such as open shortest path first (OSPF) [10] or Intermediate System to Intermediate System (IS-IS) [11], always runs and exchanges the data plane resource information such as the number of links and link bandwidth. This means that all switches have the data plane resource information of the entire network. Each ingress node of an Ethernet network can calculate a suitable path route from ingress node to egress node with algorithms such as constrained shortest path fast (CSPF). The forwarding table of the node along the calculated path route can be set up with GMPLS signaling, and nodes have VLAN forwarding tables in the data plane, as shown in Figure 2.

In global repair with protection, two disjoint paths (c.f. working and protection) are calculated and set up in Ethernet network from ingress node to egress node. When fault occurs in nodes or links on working path, node that detected fault sends messages to the ingress node, which switches from working path to protection path.

3.2. Proposed Extension for Local Repair with Protection. We propose a new fault recovery method. Nodes detecting a fault can send frames without waiting for the switching-over of conventional protection to be completed.

We introduce a new forwarding table, called "recovery table" for use when a fault occurs to achieve fast recovery. This table is created from topology information beforehand and set up with the GMPLS control plane.

The shortest path tree, whose root node is the egress node in the Ethernet network, is calculated with a shortest path calculation algorithm, such as Dikstra, and the recovery table of each node is set up. In this table, the egress node MAC address and output port are coupled. The frames, whose destination MAC address is equal to the egress node MAC address, are forwarded to this output port.

Each node has two types of forwarding tables. One is called a "forwarding table", which is usually used in the manner shown in Figure 2, and the other is called "recovery table" (Figure 3), which is used during failure.

Usually, frames, whose VLAN-ID = x, are forwarded from ingress node to egress node along with the forwarding table. When a link fault occurs and a node detects that the output port = B is not available, frames that come from input port = A and VLAN-ID = x cannot be forwarded to output port = B. This node performs global repair and local repair. For the former, the node sends the fault notification message to the ingress node for switching over from the working path to the protection path. For the latter, the node changes the VLAN-ID from x to y and forwards that frame to the output port by using the recovery table. The successive node forwards the frames by using the recovery table with the destination MAC addresses, and the frames are forwarded along the shortest path to the egress node. Therefore, nodes can forward frames without waiting for the switching over of global repair with protection to complete.

After the ingress node receives the notification message, it switches over from the working path to the protection path, the frames are forwarded through the protection path, and the frames are not forwarded to the faulted node.

In the following section, we describe the creation of a recovery table and the forwarding method of our proposed fast fault recovery method.

3.2.1. Creating Recovery Table. When a network topology is generated or changed, the GMPLS control plane determines the next hop nodes to which it should forward frames when a fault occurs.

This detection is determined by the shortest path calculation algorithm using the network topology information obtained from a routing protocol such as OSPF or IS-IS. The shortest path tree, in which the root node is the egress node, can be calculated and a recovery table created. The number of route entrees in the recovery table can be up to the number of egress nodes. Thus, network scalability will not be reduced with this mechanism.

In Figure 3, for example, it is assumed that node 2 will forward frames to the next hop node, node 6, when a failure occurs. From this output port information, a recovery table is created. Thus, when a fault occurs in the link 2-3, node 2 can forward frames to node 6 according to its recovery table.

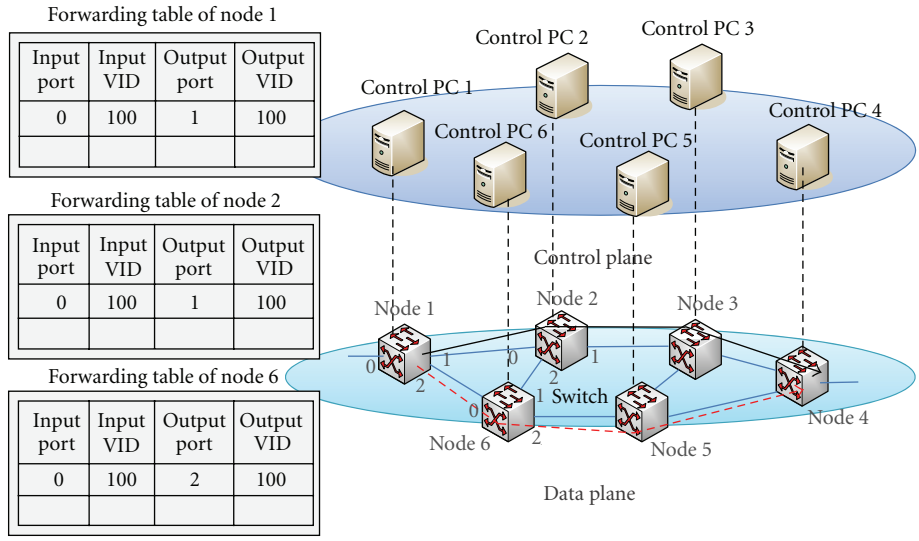

Figure 2: Path setup from ingress node to egress node.

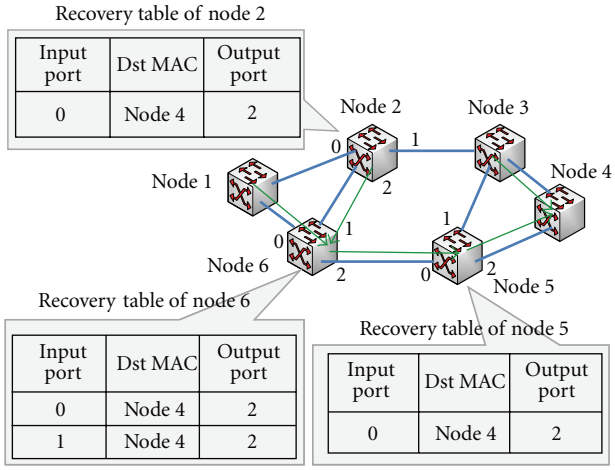

Figure 3: Proposed Recovery Table.

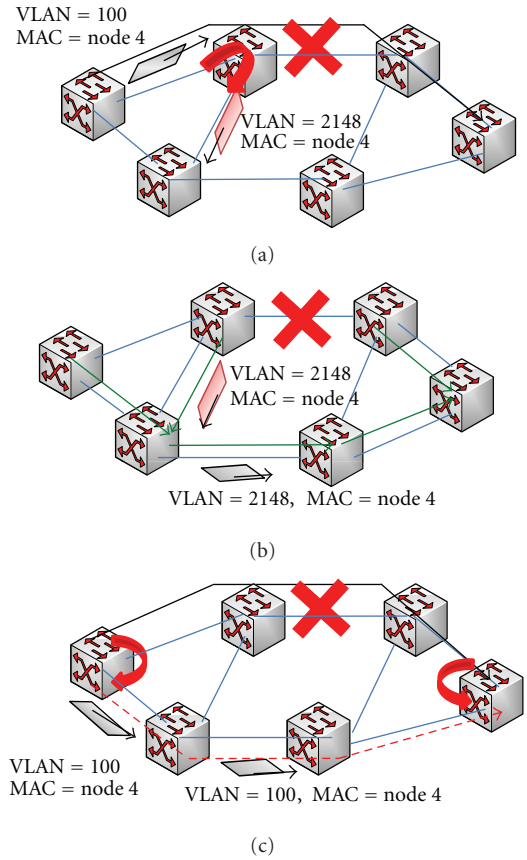

Figure 4: Forwarding action outline.

3.2.2. Forwarding Method When Fault Occurs. When a fault does not occur, nodes forward frames based on VLAN-IDs. A frame's upper one bit of VLAN-ID, which is used as a flag showing that a fault occurred, is usually set to 0.

Figure 4 shows the outline of the forwarding action when a fault occurs.

When the node that detected the fault receives the frames that should be forwarded to the failed link, it sets these frames' flags to 1, and outputs frames to the output port, which is shown in its recovery table.

When the node receives the frames whose flags are set to 1 it searches for the destination MAC address of the frames and determines the output port of these frames according to its recovery table.

However, in this method, bandwidth in the forwarding path when a fault occurs is not reserved beforehand. Therefore, when a fault occurs, temporary flows suppress the bandwidth used by normal flows in the forwarding route. A solution for this problem is described in the following section.

3.2.3. Bandwidth Control and Using Multiple Paths. To solve the problem with the proposed method in which temporary flows suppress the bandwidth used by normal flow in

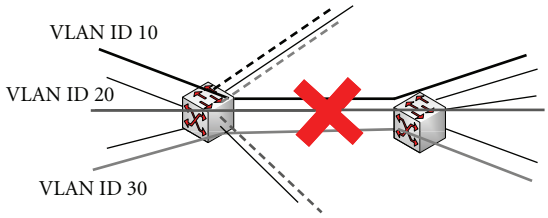

Figure 5: Bandwidth suppression by temporary flow.

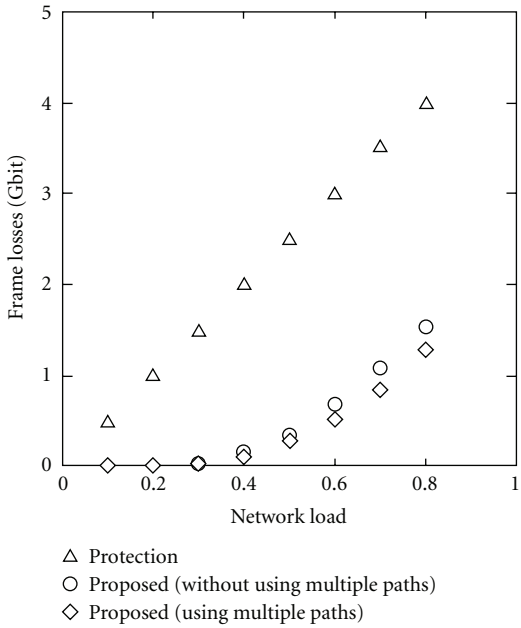

Figure 6: Relation between network load and frame losses.

the forwarding route, the bandwidth control function is implemented in the layer-2 switches. With this function, normal flow can ensure their bandwidth, and temporary flow can use only the remainder of the bandwidth.

Frames of temporary flow can be stored in the buffer of the switch. However the capacity of the buffer is limited; therefore, frames of temporary flow are lost when the buffer overflows.

By using multiple paths such as equal cost multipath (ECMP) of OSPF, when a fault occurs, frame losses of temporary flow can decrease. When a fault occurs, the node forwards frames using multiple paths based on their VLAN-IDs in a round-robin method, as shown in Figure 5. Frames whose VLAN-ID is 10 or 30 are forwarded to the upper path, and those whose VLAN-ID is 20 are forwarded to the lower path.

4. Performance Evaluation

We conducted a computer simulation to evaluate the frame losses of the proposed and conventional methods.

4.1. Assumption of Simulation. The network topology, in which the number of nodes is 100 and the average link degree

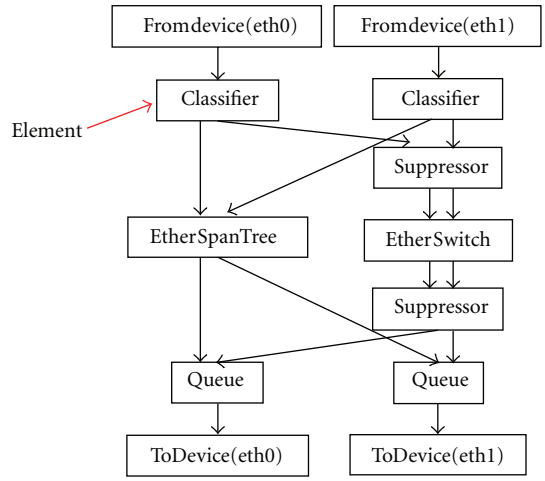

Figure 7: Structure of click.

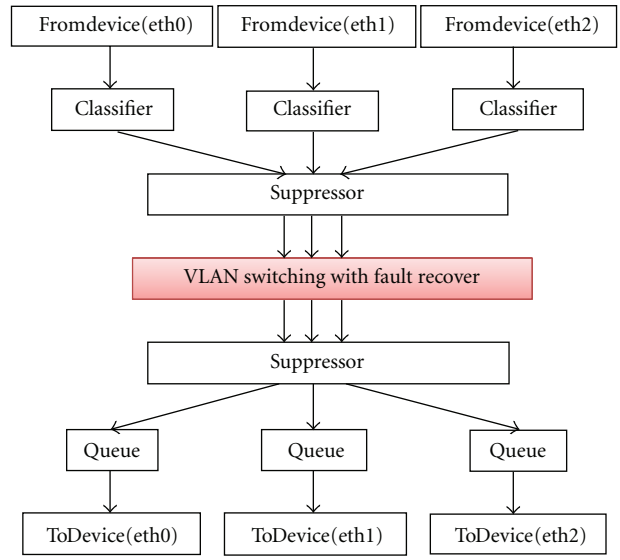

Figure 8: Structure of implemented recover function.

is 6, is randomly created. We assumed that the bandwidth of all links is 100 Gbps and the recovery time of protection is 50 ms. Frame losses were evaluated in the proposed method by using and not using multiple paths.

4.2. Simulation Results. Figure 6 shows the relation between the network load and frame losses when a fault occurs. The proposed method reduces the frame losses by more than 1 Gbits. This shows that nodes can determine the next hop nodes, regardless of network topology, and more than 1 Gbits are forwarded by using only the remainder of the bandwidth of the forwarding route. By using multiple paths, frame losses can also be decreased, especially in high network loads.

5. Experiments

In this section, we discuss the experiments on the proposed fault recovery method implemented on the software switch "Click Modular Router" [12].

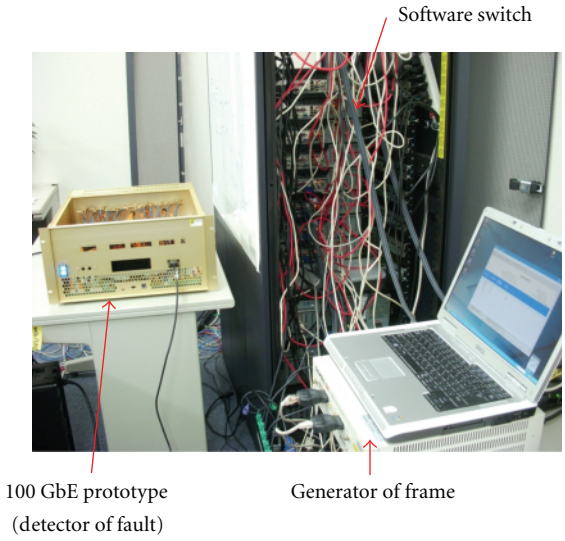

Figure 9: Experimental system.

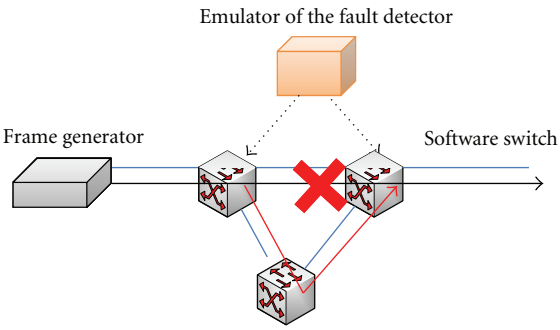

Figure 10: Experimental network.

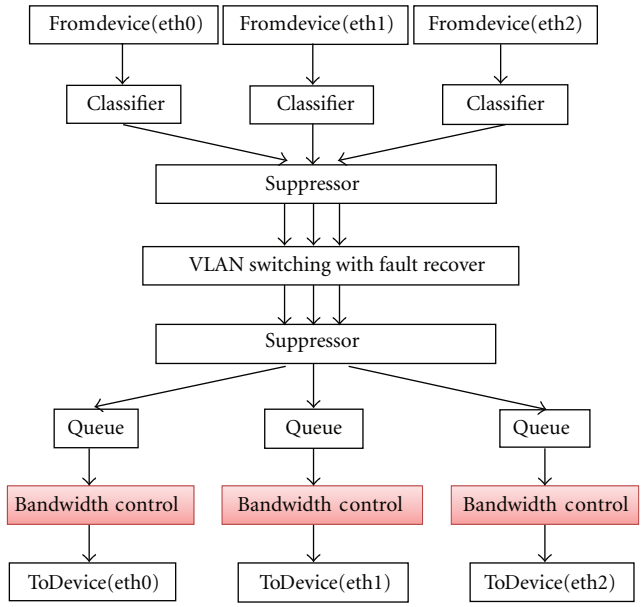

Figure 11: Implementation structure of bandwidth control.

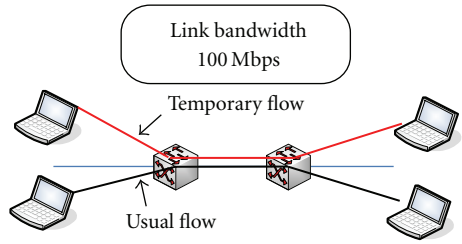

Figure 12: Experimental network.

We measured the recovery time in the proposed method when a fault occurred and evaluated the effectiveness of the bandwidth control function implemented on the layer-2 switch.

5.1. Software Switch: Click Modular Router. We now discuss the software switch "Click Modular Router," which was used to implement the proposed method.

The switch is constructed using segmented functions of the switch. Figure 7 shows an example of the structure of the Ethernet switch.

By making new elements, we can add a new function such as fault recovery, bandwidth control, or delay management to the switch.

5.2. Measurement of Recovery Time

5.2.1. Experimental Network. The structure of the implemented layer-2 switch in the software switch is shown in Figure 8. We developed the "VLAN Switching with fault recovery" element for our proposed method.

Figure 9 is a photograph of the experimental system. A 100-GbE prototype (fault detector) and frame generator were connected to the software switches. The experimental network in which we measured recovery time is shown in Figure 10.

We developed a fault-detector emulator. It sends a fault-detection signal to one switch and, after 3 μs, sends a signal to the other switch. 3 μs is the fault detection time (80 ps $\times$ 2112 bits $\times$ 16 $\fallingdotseq$ 3 μs) because the actual detector determines a channel (12.5 Gbps) as failed when it continuously detects 16 errors in the forward error correction (FEC) frame (2112 bits).

5.2.2. Results of Experiment on Proposed Fast Recovery Method. In this experiment, the transmission rate was 10 Mbps because of the performance of the software switches, and frames flowed for over 20 sec. Total data amount is a 25-MByte.

Table 1 lists the results of measured frame losses. The frame losses by fault are calculated by subtracting the frame losses before fault from those after fault. The frame losses by fault correspond to about 20 ms of link failure because the transmission rate was 10 Mbps.

This result shows that by applying the proposed method, recovery time is about 20 ms.

Table 1: Measured frame losses.

Frame size	Frame losses before fault	Frame losses after fault	Frame losses by fault
1518 Bytes	151 KBytes	178 KBytes	27 KBytes
512 Bytes	163 KBytes	193 KBytes	30 KBytes

The above results were for software. Hardware recovery time is expected to be faster.

5.3. Evaluation of Bandwidth Control Function

5.3.1. Experimental Network. The bandwidth control function, in which normal flow can ensure their bandwidth and temporary flow can use only the rest of the bandwidth, is implemented on the layer-2 switch.

The structure of the implemented layer-2 switch in software switches is shown in Figures 11 and 12 and shows the experimental network.

The link bandwidth was 100 Mbps, and the bandwidth used by normal flow was always 40 Mbps. Under this condition, the bandwidth required by temporary flow changed from 10 to 100 Mbps, and the bandwidth used by temporary and normal flows were measured.

5.3.2. Results of Experiment on Proposed Bandwidth Control Function. Figures 13 and 14 show the relation between the bandwidth required by temporary flow and that used by normal and temporary flows.

Figure 13 shows the results when the switches do not have the bandwidth control function, and Figure 14 shows the results when the switches have the bandwidth control function. If switches do not have the bandwidth control function, the bandwidth used by the normal flow decreases as the temporary flow requires more bandwidth.

If switches have the bandwidth control function, the bandwidth used by the normal flow can maintain 40 Mbps because the bandwidth used by the temporary flow is limited to 60 Mbps, that is, the remainder of the bandwidth. Therefore, the bandwidth control function implemented on the layer-2 switch is effective.

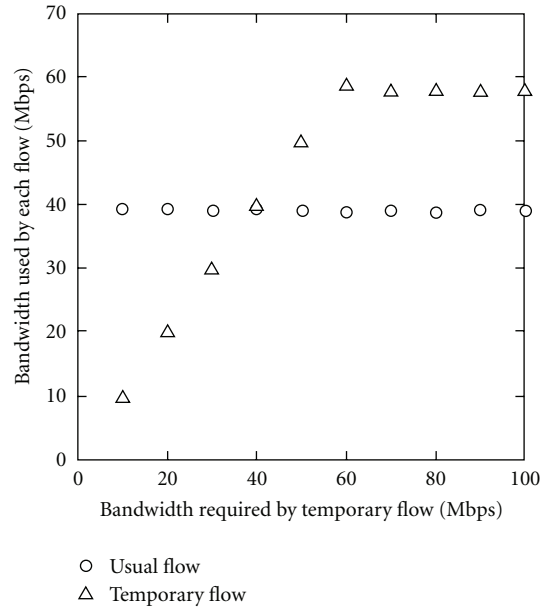

Figure 13: Bandwidth used by normal and temporary flows when bandwidth required by temporary flow changes when switches do not have bandwidth control function.

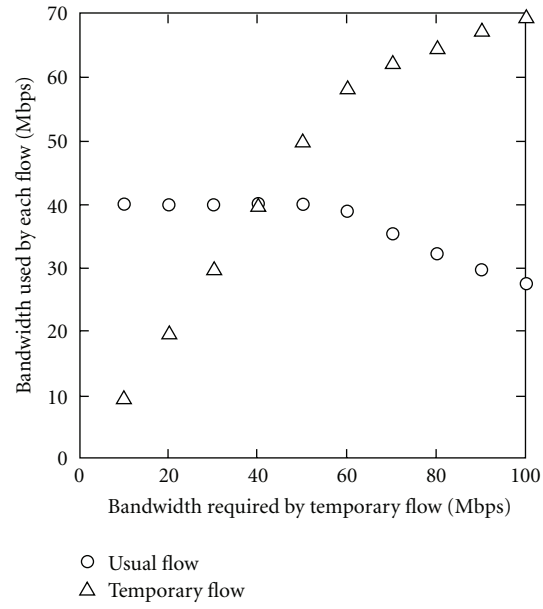

Figure 14: Bandwidth used by normal and temporary flows when bandwidth required by temporary flow changes when switches have bandwidth control function.

6. Conclusion

We proposed a new fault recovery method with a preestablished recovery table, in which nodes can send frames without waiting for the switchingover of protection to complete.

Each node makes its recovery table from topology information when a network topology is changed. When a fault occurs, nodes can forward frames by searching its recovery table.

The computer simulation showed that by applying the proposed method, frame losses can decrease by more than 1 Gbit. Moreover, by using multiple paths when a fault occurs, frame losses can decrease.

The experimental results showed that by applying the proposed method, recovery time is about 20 ms, even in software. In hardware, fault recovery is expected to be faster.

We also proposed a bandwidth control function and implemented it in the layer-2 switch which is also effective for fast fault recovery. Therefore, our proposed fast fault recovery method can create a reliable wide area layer-2 network.

Acknowledgments

This work was partially supported by "Lambda Access" Project funded by the National Institute of Information and Communication Technology (NICT). This work was also supported by the Japan Society for the Promotion of Science's (JSPS) Grant-in-Aid for Scientific Research (19360178) and by a Grant-in-Aid for the Global Center of Excellence for high-Level Global Cooperation for Leading-Edge Platform on Access Spaces from the Ministry of Education, Culture, Sport, Science, and Technology in Japan. I'm grateful to HITACHI for offering 100 GbE prototype.

References

[1] K. Fouli and M. Maier, "The road to carrier-grade ethernet," *IEEE Communications Magazine*, vol. 47, no. 3, pp. S30–S38, 2009.

[2] IEEE Std 802.1Q, "IEEE Standards for Local and Metropolitan Area Networks Virtual Bridged Local Area Network," 2006, http://standards.ieee.org/getieee802/download/802.1Q-2005.pdf.

[3] G. Chiruvolu, A. Ge, D. Elie-Dit-Cosaque, M. Ali, and J. Rouyer, "Issues and approaches on extending Ethernet beyond LANs," *IEEE Communications Magazine*, vol. 42, no. 3, pp. 80–86, 2004.

[4] M. Ali, G. Chiruvolu, and A. Ge, "Traffic engineering in metro ethernet," *IEEE Network*, vol. 19, no. 2, pp. 10–17, 2005.

[5] E. Mannie, Ed., "Generalized Multi-Protocol Label Switching (GMPLS) Architecture," IETF RFC 3945, 2004.

[6] R. Aubin and H. Nasrallah, "MPLS fast reroute and optical mesh protection: a comparative analysis of the capacity required for packet link protection," in *Proceedings of the Design of Reliable Communication Networks (DRCN '03)*, pp. 349–355, Alberta, Canada, October 2003.

[7] M. Batayneh, B. Mukherjee, D. A. Schupke, M. Hoffmann, and A. Kirstaedter, "Carrier-grade ethernet: etherpath protection vs. Ethertunnel protection," *IEEE Network*, vol. 23, no. 3, pp. 10–17, 2009.

[8] S. Ramamurthy and B. Mukherjee, "Survivable WDM mesh networks. Part I-protection," in *Proceedings of the 18th Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM '99)*, pp. 744–751, 2, March 1999.

[9] S. Ramamurthy and B. Mukherjee, "Survivable WDM mesh networks. Part II-restoration," in *Proceedings of the 18th Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM '99)*, vol. 2, pp. 2023–2030, March 1999.

[10] D. Katz, G. Xiao, and T. H. Cheng, "Traffic Engineering (TE) Extensions to OSPF Version2," RFC3630, 2003.

[11] D. Oran, "OSI IS-IS Intra-domain Routing Protocol," RFC 1142, 1990.

[12] E. Kohler, R. Morris, B. Chen, J. Jannotti, and M. F. Kaashoek, "The click modular router," *ACM Transactions on Computer Systems*, vol. 18, no. 3, pp. 263–297, 2000.