A Network-disaster Recovery System using Multiple-backup Operation Planes

Toshiaki Suzuki, Hideki Endo, Isao Shimokawa,
and Kenichi Sakamoto
Research & Development Group
Hitachi, Ltd.
Kanagawa, Japan
{toshiaki.suzuki.cs, hideki.endo.es, isao.shimokawa.sd,
and kenichi.sakamoto.xj}@hitachi.com

Hidenori Inouchi, Taro Ogawa, Takanori Kato,
and Akihiko Takase
Information & Telecommunication Systems Company
Hitachi, Ltd.
Kanagawa, Japan
{hidenori.inouchi.dw, taro.ogawa.tg, takanori.kato.bq,
and akihiko.takase.wa}@hitachi.com

***Abstract*—A "network-disaster recovery system" using multiple-backup operation planes is proposed. Under this system, a whole network is separated into multiple areas. Before starting network operations, a network-management server calculates recovery paths for every possible failure in network area and distributes them with a recovery identifier (ID) for each network-area-failure pattern (on the backup operation plane). Network nodes receive and store the recovery IDs and recovery configurations. The network-management server determines a failure pattern after detecting the network-area failures and distributes the recovery ID to related network nodes. The network nodes that received the recovery ID start data transmission according to the path configurations specified by the recovery ID. After the completion of these procedures, the network-area failures are swiftly recovered. A prototype system (composed of a network-management server and 96 simulated packet-transport nodes) with a graphical viewer was implemented, and its performance was evaluated. According to the results of the evaluation, all recovery-path configurations for 1000 pseudo-wires (PWs) (namely, transmitting the recovery ID to the related network nodes and using a recovery-path database specified by the ID) were done within 100 milliseconds after the network-area failures were detected. On the condition that the configuration time depends on the size of the recovery-path database, the proposed system takes about one minute and 40 seconds in the case of 1,000,000 PWs. On the other hand, a restoration scheme under the same evaluation conditions used for the proposed system takes over 10 minutes to recalculate recovery paths from detection of the first area-based network failure. That is, the proposed recovery scheme can recover network-area failures faster than the conventional restoration scheme can.**

Keywords - network management; disaster recovery; packet transport; reliable network

I. INTRODUCTION

Lately, as reflected in the rising number of Internet users and the popularity of cloud services, applications and services provided by way of networks have become indispensable in daily life. Network services must, therefore, be highly reliable and "always available". When extensive disasters occur, network services could be out of service for a long time. Consequently, networks must be robust enough so that they can continue to provide services even if their facilities are extensively damaged. In our previous study, presented at INNOV 2014 [1], entire system architecture was focused on. In this extended work, a prototype system was implemented, and its performance was evaluated in comparison with a conventional system.

As recovery procedures for network failures, two major techniques [2] are applied: "protection," by which recovery paths are physically prepared in advance of network failures by allocating extra network resources; and "restoration," by which recovery paths are "calculated" after network failures are detected.

Protection is easily applied to multi-layer networks, and recovery is immediate because recovery paths are prepared in advance (that is, before network operations are started). However, if the prepared recovery paths are not available when network failures occur, network-connection services will become out of service. On the other hand, if restoration is applied, network connections can be recovered if recovery paths are recalculated after network failures are detected. However, it takes more time to recalculate the recovery paths if the operated networks are huge and have many network nodes. Therefore, if huge quantities of paths are used to transmit data packets, much time is needed to recalculate all recovery paths, and the network will not recover from a disaster expeditiously. In addition, even if network connections are recovered, all network flows will try to use the same recovery path. As a result, the network will easily become congested, making it difficult to guarantee network-transmission quality.

In light of the above-described issues, a robust network-management scheme is required. Specifically, it must control multi-layer-network resources so as to provide and maintain network-connection services at times of a "network disaster" (namely, a catastrophic failure of a network). To achieve that control, a network-management system has to monitor and control the multi-layer-network resources.

The overall aim of the present study is to develop a network-management scheme for monitoring and controlling multi-layer network resources so as to provide robust networks that can swiftly recover from a network disaster. To swiftly recovery from a network disaster, three steps should be followed: the first step is to find network failures in a short time; the second is to promptly determine how to recover the network; and the third is to immediately configure recovery paths. In the present study, the second step is focused on, and a "network-disaster recovery system" using an area-based network-management scheme, which controls networks composed of IP networks and packet-transport networks, such as the Multi Protocol Label Switching - Transport Profile (MPLS-TP) network, is proposed.

The rest of this paper is organized as follows: Section II describes related work. Section III explains the requirements concerning a network-disaster recovery system. Section IV proposes a network-disaster recovery system that meets those requirements. Sections V and VI respectively describe an implementation of a prototype system and present some results of evaluations of the system's performance. Section VII concludes the paper.

II. Related Work

Regarding highly available and reliable network management, several standardization activities have been ongoing. For example, MPLS-TP-related operation, administration, and maintenance (OAM) functions have been standardized. In the first stage of that standardization, the International Telecommunication Union - Telecommunication Standardization Sector (ITU-T) [3] discussed specifications such as Transport – Multi Protocol Label Switching (T-MPLS). In the next stage, the ITU-T jointly standardized MPLS-TP specifications with the Internet Engineering Task Force (IETF) [4]. A request for comments (RFC) on requirements of MPLS-TP [5] was issued as the first step. A framework of MPLS-TP was documented as RFC 5921 [6]. Using MPLS-TP OAM [7] functions makes it easier to detect failures in transport networks. In addition, an RFC on a framework for MPLS-TP survivability [8] was issued. In relation to the proposed system, it is useful to detect network failures promptly in order to determine areas that are out-of-service.

With regards to failure recovery, two major techniques, namely, "protection" and "restoration," have been proposed. By means of protection, a standby path is preliminarily calculated and established by using extra physical resources. When network failures are detected, an active path is promptly changed from the current path to the standby path.

One of major recovery schemes, called "fast reroute" [9], prepares a back-up path. In addition, a recovery scheme combining an IP layer and an optical layer was proposed [10]. A scheme for preparing multiple backup paths to tackle multiple failures was presented [11]. A recovery procedure for multiple levels [12], such as global, segment, and local protection, was studied. A network-protection scheme for guaranteeing recovery time [13] was also proposed. In addition, a protection mechanism using fewer network resources by sharing wavelength-division-multiplexing (WDM) resources [14] was issued. By means of this protection mechanism, in the case of multiple network failures, a large number of standby paths are prepared, so a huge volume of physical resources might be needed. It is therefore only useful for limited network failures, such as failures of a few links or nodes.

On the other hand, by means of restoration, recovery paths are calculated one by one after network failures are detected. Restoration schemes for handling multiple failures [15], considering global and local wavelength availability [16], and for virtual networks [17] have also been proposed. In addition, a fast connection-recovery scheme that reduces the search range by using special nodes as a landmark [18] was proposed. This scheme is useful for catastrophic network failures, since all reroutes are basically calculated after the failures are detected. However, if there are a large number of current paths, it might take much time to calculate all recovery paths to the current paths.

III. Requirements concerning a Network-Disaster Recovery System

The target network structure is shown conceptually in Figure 1. It is composed of an IP network layer and a packet-transport-node (PTN) network layer, such as an MPLS-TP network, controlled by a network-management server (assumed to be connected to all PTNs). The core network is composed of PTNs, while the access network is composed of IP network nodes. In this study, recovery from multiple network failures on the IP and PTN networks (for example, the two network failures shown in the figure) is focused on in this study. One of the critical issues concerning network recovery in the case of a network disaster is the time taken to recover numerous established paths of a packet network. At that time, each path is configured by a label-switched path (LSP) [19] and a pseudo-wire (PW) [20]. Specifically, the main issue is the time taken to recalculate numerous recovery paths one by one after disconnected paths are detected by monitoring network conditions.

In the case of a packet-transport network, the bandwidth of a network path is guaranteed. Therefore, ensuring the quality of a recovery path, such as bandwidth and/or end-to-end delays before (as well as after) a network failure is also an issue.

To tackle the above-mentioned issues, the proposed network-disaster recovery system should satisfy the following four requirements.

① Manage multi-layer networks

② Recover from multiple network failures

③ Rapidly establish recovery paths

④ Guarantee quality of recovery paths after network failures are recovered

To meet these requirements, the network-disaster recovery system is designed on the basis of the following policy: If plenty of paths are set up, recovery paths should not be recalculated after multiple network failures are detected (since it takes considerable time to recalculate them). On the other hand, recovery paths that guarantee bandwidths and delays for each possible network failure should be calculated preliminarily, and paths should be promptly recovered by using the prepared paths after the network failures are detected.

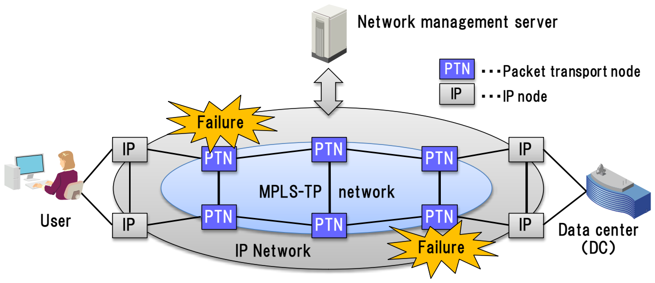

Figure 1. Target network structure

IV. Proposed Network-disaster Recovery System

As for the proposed network-disaster recovery system, a network-management server centrally manages an entire network. In the target network, a core-network segment is composed of PTNs, and an access-network segment is composed of IP network nodes. In addition, the network-management server manages the entire network by dividing it into multiple network areas and controlling each of them by using the area-based network-management scheme.

A. Structure of proposed system

The structure of the proposed network-disaster recovery system is shown in Figure 2. As an example of area-based management, the network-management server divides the whole PTN network into eight areas by using a conventional scheme, such as cluster analysis, and manages them by using an area-based management scheme. The eight areas are shown as network (NW) areas (1) to (8) in the figure. In addition, the network-management server is assumed to be connected to all PTNs, a user terminal, and servers in a datacenter (DC) by another management network (not shown in the figure). The network-management server monitors all PTNs, manages available network resources, and keeps them as topology-related data. It also executes swift network-disaster recoveries after detecting catastrophic network failures.

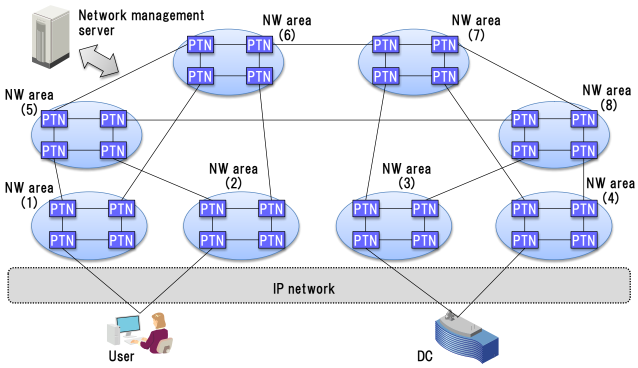

Figure 2. Proposed network-disaster recovery system

As for the proposed disaster-recovery system, the user terminal is connected to a server in the DC via the IP network and PTN networks, and it can get various application services from the server. To provide the user terminal with robust network access, it is connected to at least two "PTN network areas". In addition, the DC is connected to at least two other "PTN network areas".

B. Overview of network-disaster recovery system

The two main procedures used by the proposed network-disaster recovery system are overviewed in Figure 3. Following the first procedure, the network-management server divides an entire PTN network into eight network (NW) areas, labelled (1) to (8) in the figure, and controls them by using the area-based network-management scheme. In addition, it configures the path shown as the solid red lines in the figure as the current path so that the user can access the server in the DC and use application services.

Following the second procedure, the network-management server preliminarily calculates all recovery paths by considering all possible area-based failures. Specifically, the number of possible area-based failure patterns is 255 (since there are eight areas, and each area could be independently active or non-active), namely, 256 (i.e., 2^8) patterns minus a "no network area failure" pattern that is the current network operation. The network-management server assigns a recovery ID for each area-based network failure pattern and stores each recovery ID with information about the recovery paths. It then preliminarily distributes all recovery IDs and the recovery-path information to all PTNs. As stated in Figure 3, it is assumed that network areas (1), (3), and (6) fail. In the case of these failures, the path depicted by a dashed line is prepared as a recovery path, and the recovery-path information is distributed to PTNs related to that recovery path before network operations are started.

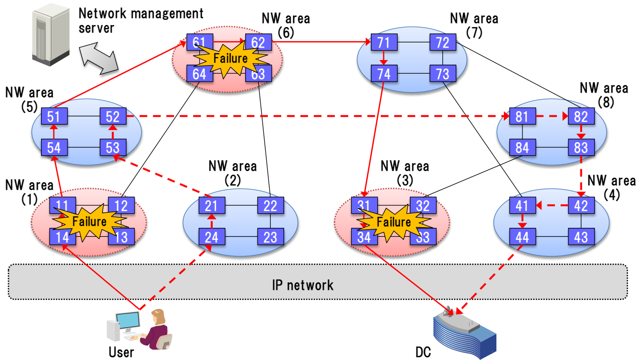

Figure 3. Proposed procedures for network-disaster recovery

During network operations, the network-management server monitors area-based network failures. When it detects an area-based network failure, it determines a failure pattern and a recovery ID. It then distributes the recovery ID to related PTNs, the user terminal, and the server in the DC. The PTNs that receive the recovery ID start to promptly recover and transmit packet data according to the recovery-path information specified by the ID. In addition, the network-management server configures IP networks to transmit packet data from the user terminal to network area (2). Alternately, if necessary, it transmits a request that asks the user terminal to change the output port so as to transmit packet data to another active network area. Besides, the network-management server configures IP networks to transmit packet data from network area (4) to the server in the DC.

C. *Sequence of network-disaster recovery*

The proposed network-disaster recovery follows the sequence shown in Figure 4. First, the network-management server divides the entire PTN network into multiple network areas and manages each area by using the area-based network-management scheme, labeled “area mgmt” in Figure 4. Specifically, the PTN networks are divided into eight areas and managed as shown in Figure 3. Subsequently, the network-management server calculates and configures a path as the current path (which is composed of a LSP and a PW) for transmitting packet data from the user terminal to the server in the DC, shown as “current-path configuration” in Figure 4. It starts network operations after configuring the calculated path to related PTNs. As for calculating a path, a route that can provide required bandwidths and transmit packet data within allowed delays is selected as the current path.

The network-management server then calculates and configures all recovery paths, shown as “recovery-path configuration” in Figure 4, by considering all possible area-based network failures. The recovery paths are calculated by a conventional scheme, such as Dijkstra’s algorithm, considering remaining network resources as a backup operation plane for each possible area-based network failure, as shown in Table I. The “recovery-path configuration” shown in the table provides a list of nodes through which data transit. Available network resources are managed by excluding resources belonging to an assumed failure area. Specifically, each recovery path (labeled “P1” in the table) is identified by a recovery ID from “0” to “255.” The top row of the table, containing recovery ID “0”, indicates current-recovery-path configurations for no area-based network failures.

The next row in the table, containing recovery ID “1”, indicates recovery-path configurations for recovering a failure of network area (1). In this case, it is assumed that the network failure occurs in area (1). The recovery path “P1” is calculated on the basis of available network resources. In other words, network resources in area (1) are excluded from the available resources, and the recovery path is calculated. The next row in the table, containing recovery ID “2”, indicates the recovery-path configurations for recovering a failure of network area (2). The row containing recovery ID “38” indicates the recovery-path configurations in the case of failures of network areas (1), (3), and (6). As an example recovery path, the dashed line in Figure 3 is that for the current path depicted by the solid line. In Figure 3 and Table I, only the recovery-path information for path “P1” is shown as an example. However, the proposed system can manage multiple paths.

As the next step of a recovery, the network-management server calculates recovery-path configurations for each node in each area-based network-failure pattern according to the recovery-path information shown in Table I. As examples, the recovery-path configurations for PTN 53 and 54 are listed in Table II. The information in Table II is obtained by restructuring the node lists of the recovery-path configuration shown in Table I. For example, if PTN 53 is focused in, nodes that are connected to PTN 53 are gathered from Table I and sorted as shown in Table II.

The top row of the table, containing recovery ID “0” on PTN 53, shows the current configuration (i.e., “connection 1” and “connection 2”). With regard to PTN 53, the path P1 (composed of an LSP and a PW) is not configured, since it does not transmit the related packet data. The next row of the table, containing recovery ID “1”, indicates the configuration for recovery path P1 in the case of a failure of network area (1). Specifically, it is shown that PTN 53 transmits packet data of P1 from PTN 21 to PTN 54 and from PTN 54 to PTN 21. In addition, the row of the table containing recovery ID “2” indicates the recovery-path configurations in the case of a failure of network area (2). In this case, the recovery-path configurations for P1 are not included, since PTN 53 does not transmit data for P1. On the other hand, the row of the table containing recovery ID “38” indicates the configurations of recovery path P1 in the case of failures of network areas (1), (3), and (6). Specifically, it is shown that PTN 53 transmits packet data of path P1 from PTN 21 to PTN 52 and from PTN 52 to PTN 21.

In the lower half of the table, recovery-path configurations for PTN 54 are indicated. The row of the table containing recovery ID “0” shows the current configuration. As shown in the table, PTN 54 transmits packet data of path P1 from PTN 11 to PTN 51 and from PTN 51 to PTN 11. The row of the table containing recovery ID “2” indicates the recovery-path configurations in the case of a failure of network area (2). PTN 54 transmits packet data of path 1 from PTN 11 to PTN 51 and from PTN 51 to PTN 11. In addition, the row of the table containing recovery ID “38” indicates the recovery-path configuration in the case of failures of network areas (1), (3), and (6). However, recovery path 1 is not configured, since PTN 54 does not transmit path-1-related packet data. After the network-management server calculates all recovery-path configurations shown in Table II, it distributes them to all PTNs. When each PTN receives the configurations, it stores them with each recovery ID.

In the next step of the recovery sequence, the network-management server monitors operations of all PTNs and area-based network failures, shown as “monitoring” in Figure 4. For example, the network-management server detects the failures of network areas (1), (3), and (6) shown in Figure 3. In this case, the network-management server selects recovery ID 38 to recover the configured path, shown as "recovery decision". The PTNs receive recovery ID 38 and configure a data-transmission function to transmit packet data according to the recovery-path information specified by recovery ID 38, shown as "recovery ID distribution".

In the next step, the network-management server configures IP networks to transmit packet data from the user terminal to PTN 24. In addition, it configures IP networks to transmit packet data from PTN 44 to the server in the DC, shown as "recovery configuration". In summary, executing the above-described recovery procedures makes it possible to recover failures of network areas (1), (3), and (6).

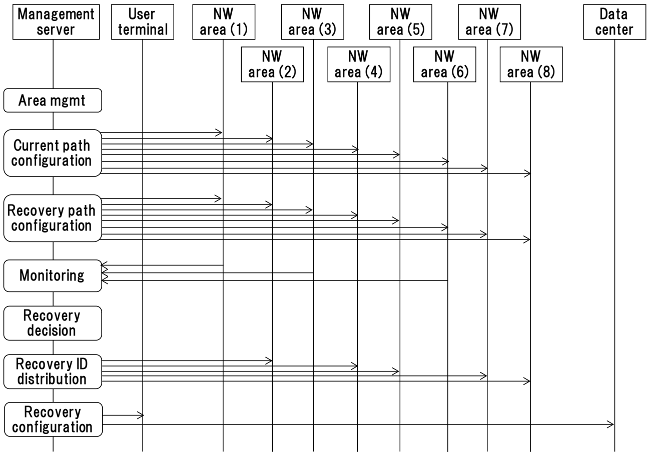

Figure 4. Sequence of network-disaster recovery

TABLE I. RECOVERY-PATH CONFIGURATIONS

Failure pattern	Recovery ID (operation plane)	Path	Recovery path configuration
No failure	0	P1	14, 11, 54, 51, 61, 62, 71, 74, 31, 34
Area (1) failure	1	P1	24, 21, 53, 54, 51, 61, 62, 71, 74, 31, 34
Area (2) failure	2	P1	14, 11, 54, 51, 61, 62, 71, 74, 31, 34
- - -	- - -	- - -	- - -
Areas (1), (3), (6) failures	38	P1	24, 21, 53, 52, 81, 82, 83, 42, 41, 44
- - -	- - -	- - -	- - -
All-area failures	255	P1	No recovery

TABLE II. RECOVERY-PATH CONFIGURATIONS FOR EACH PTN

PTN	Recovery ID	Path (LSP/PW)	Connection 1	Connection 2
53	0	- - -	- - -	- - -
	1	P1	21	54
	2	- - -	- - -	- - -
	- - -	- - -	- - -	- - -
	38	P1	21	52
	- - -	- - -	- - -	- - -
54	0	P1	11	51
	1	- - -	- - -	- - -
	2	P1	11	51
	- - -	- - -	- - -	- - -
	38	- - -	- - -	- - -
	- - -	- - -	- - -	- - -

D. *Calculation of recovery paths for possible failure patterns*

The flow for calculating a recovery path for an area-based network failure is shown in Figure 5. After the recovery-path calculation starts, delays and available bandwidths between PTNs are calculated from a database that includes topology information and available resources, such as link bandwidths. Next, a possible area-based network failure, for example, a failure of network area (1), is assumed. After that, the PTNs belonging to the assumed network-area failure are excluded from the available resources for calculating recovery paths. After available resources, such as PTNs and bandwidth, are fixed, one of the established PWs is selected to prepare a recovery path. Then, the minimum-delay path that has the same starting and ending points is selected as the recovery path (which is calculated in consideration of available bandwidth and delay). If a recovery path is not found, because of problems like link disconnection, a message indicating "lack of resources" for finding the recovery path is displayed, and the recovery-path calculation process moves on to the next step, namely, selection of another PW. If a recovery path is found, whether it meets the allowed delay time or not is checked. If the path does not meet the allowed delay time, a "lack of available resources" message is displayed, and the calculation process moves on to the next step, namely, finding a recovery path for another PW. If the path meets the allowed delay time, it is determined as the proper recovery path.

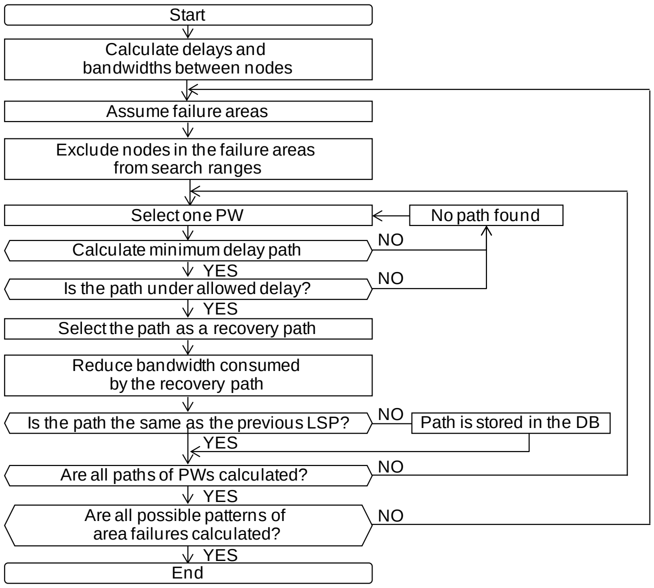
Start
Calculate delays and bandwidths between nodes
Assume failure areas
Exclude nodes in the failure areas from search ranges
Select one PW
No path found
Calculate minimum delay path
NO
YES
Is the path under allowed delay?
NO
YES
Select the path as a recovery path
Reduce bandwidth consumed by the recovery path
Is the path the same as the previous LSP?
NO
Path is stored in the DB
YES
Are all paths of PWs calculated?
NO
YES
Are all possible patterns of area failures calculated?
NO
YES
End

Figure 5. Calculation of recovery paths

After the recovery path is confirmed, available bandwidth is decreased by the amount of bandwidth consumed by the recovery path itself. Subsequently, if the route of the LSP path is not the same as the previously calculated route, it is stored as a new LSP route. Then, whether all recovery paths for a selected area-based network-failure pattern have been calculated is checked. If all recovery paths are not calculated, the process moves on to the next step, that is, selection of another PW. If all the recovery paths for one area-based network-failure pattern have been calculated, whether all recovery paths for all possible area-based network-failure patterns have been calculated is checked. When all the recovery paths for all possible area-based network failure patterns are calculated, the recovery-path calculation process stops. All recovery paths are calculated, and the recovery-path information is distributed to all network nodes, before network operations are started. Therefore, the nodes can select an appropriate recovery path swiftly when a network fails.

V. Implementation

A prototype system based on the above-described architecture was implemented by using three servers. The configuration of the prototype system—consisting of an application server, a control server, and MPLS-TP simulator server—is shown in Figure 6. Only the structure of the implemented software components is shown in the figure.

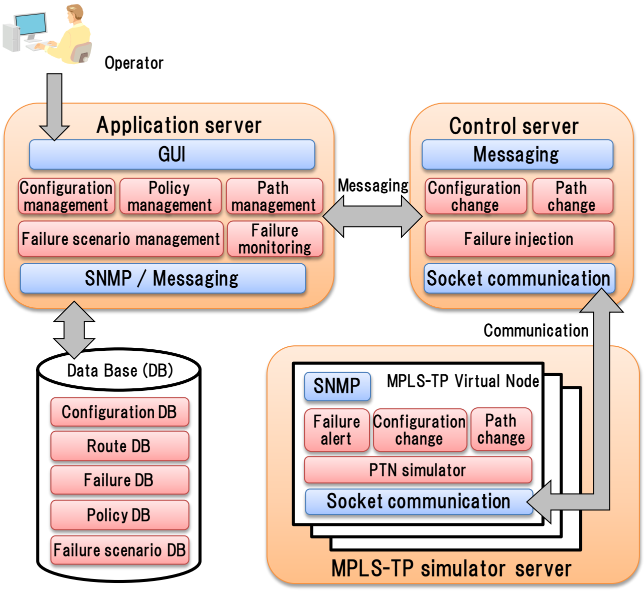

Figure 6. Structure of prototype system

The application server is in charge of overall network management. Specifically, it manages network configurations and policies so that it can select a data-transmission route. That is, it calculates the best data-transmission path for each application. In addition, it manages scenarios that would trigger network failures. The control server receives a command message from the application server and transmits it to multiple MPLS-TP virtual nodes. Specifically, it controls the network configuration and data-transmission paths. In addition, it injects network failures according to a failure request from the application server. The MPLS-TP simulator server simulates certain parts of the MPLS-TP node functions, such as changing the network configuration and setting the LSP and PW paths for data transmission. (Note that it does not simulate real data transmission.) It also detects network failures and transmits alerts to the application server.

A. *Recovery procedures executed by the prototype system*

The recovery procedures, starting with detecting alerts and finishing with recovering paths, are shown in Figure 7. First, the application server monitors network conditions. When it receives alerts of network-area failures, it analyzes them and updates the network-condition tables. In addition, it updates an alert-history table and indicates the alert on a viewer. It also analyzes the areas in which the failures occurred. The application server then determines whether a network-area failure has occurred. In the prototype system, when all PTNs in an area that receives and transmits data to other areas are damaged, the area is regarded as a being in a state of "area failure." When the application server recognizes several area failures, it determines an area-failure pattern. It then evaluates whether recovery procedures are needed. If no LSP or PW paths are damaged, even if there are area failures, recovery procedures are not taken. If there are damaged paths, recovery procedures are taken. After determining the area-failure pattern, the application server identifies a recovery ID for executing recovery procedures and transmits it to related PTNs. The PTNs change LSP and PW paths according to the path configurations specified by the recovery ID. The records of the LSP and PW paths are updated, and network-area failures are indicated on the viewer.

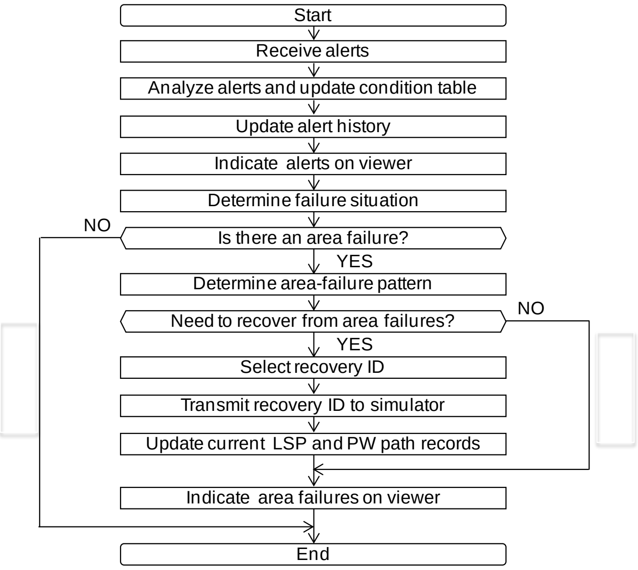
Start
Receive alerts
Analyze alerts and update condition table
Update alert history
Indicate alerts on viewer
Determine failure situation
NO
Is there an area failure?
YES
Determine area-failure pattern
Need to recover from area failures?
NO
YES
Select recovery ID
Transmit recovery ID to simulator
Update current LSP and PW path records
Indicate area failures on viewer
End

Figure 7. Recovery procedures

B. Implementation of viewer

Viewer functions that enable a user to easily understand operating conditions of the implemented network-disaster recovery system are described in this section.

1) Structure of primary screen

The structure of the primary screen of the prototype system's viewer is shown in Figure 8. The header panel includes a function menu, a user name, a logout button, and so forth. The condition panel displays the current situation regarding networks in certain areas. The topology tree shows a list of connected network nodes in a tree structure. The alert panel indicates up-to-date alerts, showing the level of severity in different colors. The map panel shows the position of the displayed network on the condition panel in relation to the entire network. A network operator can select one of the network-management functions, namely, monitoring network failures, displaying LSP and PW lists, displaying log data of failure histories, setting configurations, and setting failure scenarios, from the pull-down function menu. The "user name" tag shows the name of the current login user. The "logout" button is used to logout of a network-management function.

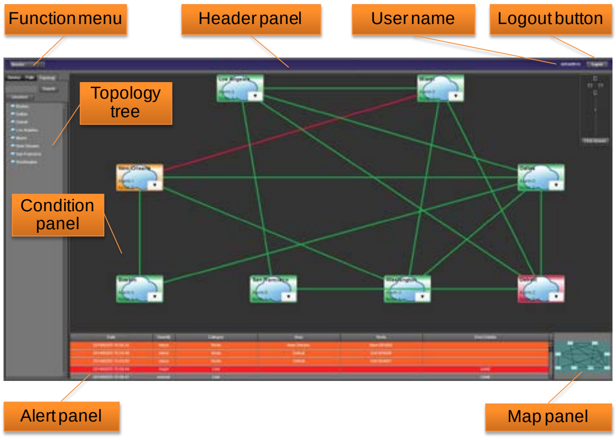

Figure 8. View of primary-screen layout

2) Large-scale-area view

A view of a large-scale area is shown in Figure 9. The "area object" tag shows an area that includes multiple PTNs. In addition, the condition of the PTNs in that area is depicted in different colors. If an area failure has not occurred, the area object is depicted in green. If several PTNs fail, they are depicted in yellow. When an area failure occurs, it is depicted in red. The "area name" tag means the name of the area. The "number of failure nodes" tag means the number of failure PTNs and is written in four digits. On the other hand, the "the number of nodes" tag means the number of all PTNs and is written in four digits. The "area enlarge" button provides a function to show the network topology of the area by a single click. The "link object" tag shows the existence of a link between areas or between an area and a user. The color of the link is depicted according to one of the following conditions: no failure, partial failure, and full failure. The "zoom palette" button provides various magnifications for viewing the displayed network area. The "scroll" button changes the displayed network area in the direction of a selected button. The "zoom bar" button rescales the displayed network area. The "automatic size control" button resizes the displayed area to its original position.

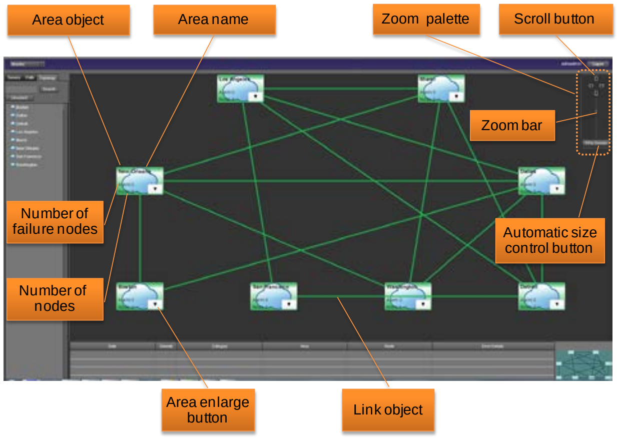

Figure 9. Large-scale area view

3) View of user-connected topology

A view of a user-connected topology is shown in Figure 10. The "list of users" tag shows users connected to the network. The "user object" tag shows an individual user. It includes a "user name" tag showing the name of the current user and a "service name" tag showing the name of the service selected by the user. The "current path highlight" tag indicates the current LSP and PW paths used for communication between users by showing multiple color-dotted lines. When the "user object" button is clicked, the current path is highlighted for a few seconds.

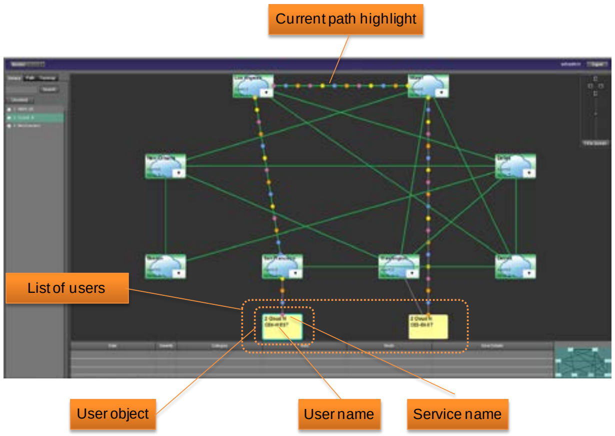

Figure 10. View of user-connected topology

4) Relation between PW and LSP

The relation between a PW and a LSP is shown in the screen view shown in Figure 11. The "user object" tag shows a connection between a user terminal and a PW path. The "PW path" tag shows relations between the user terminal and the LSP path. Each PW path has a unique name. The "edge node" tag indicates a PW edge and is connected to the user and the LSP path edge. On the other hand, the "LSP path" tag shows how the path is structured. Specifically, all PTNs that construct the LSP path are listed. Each PTN object has its own name, and the name of the area that the PTN belongs to is shown in the object. The LSP path also has a unique name, such as its number. The LSP layer is closed by pushing the "close" button.

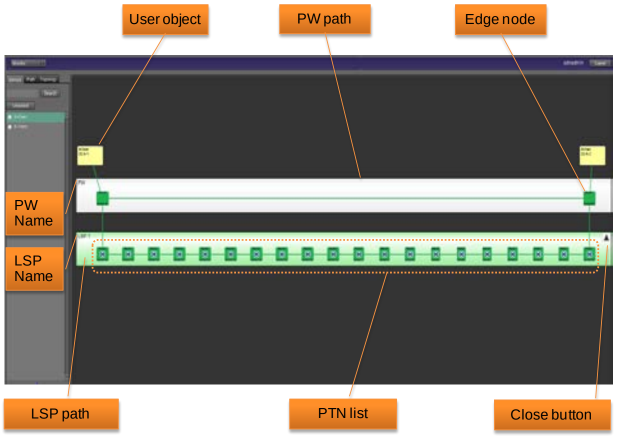

Figure 11. Relation between PW and LSP

VI. Performance Evaluation and Results

The above-described recovery procedures were evaluated in the case of multiple area-based network failures in networks composed of IP and PTN networks. First, current paths composed of LSPs and PWs were configured to allow users to access application servers in the DCs and use applications provided by the servers. In the evaluation, the procedure for recovering from multiple area-based network failures by using recovery paths was evaluated in terms of whether users can access the application servers. In addition, the time for calculating the current recovery paths and distributing the information concerning the calculated paths to all PTNs was evaluated by changing the numbers of LSPs and PWs used to construct the current paths. Specifically, the case of one user was evaluated in a previous work [1]. In this work, the case of two users was evaluated. In addition, the number of times taken in configuring a recovery path after detecting an area-base network failure was evaluated.

A. Evaluation system

The system used for evaluating the proposed recovery procedures is depicted in Figure 12. It is composed of a network-management server, PTNs, user terminals, and application servers in DCs. An entire PTN network is divided into eight network areas. Each network area is composed of 12 PTNs, as shown in NW area (7), which is an example network composed of about 100 network nodes. These PTNs are connected in a reticular pattern of 96 PTNs in total. In addition, each user terminal is directly connected to PTN-network areas (1) and (2) by the IP network, and each application server is also connected to PTN-network areas (3) and (4) directly by the IP network.

Note that the PTN networks (composed of 96 PTNs) are simulated by a physical server. In addition, the user terminal and application servers in the DCs are simulated by the same physical server. The specification of the physical server that simulates the PTN networks, user terminals, and application servers is listed in Table III. In addition, another physical server executes the network-management function, but it has the same specifications as the simulator server. In this evaluation, a system composed of eight areas and 96 PTNs is selected since it is large enough to establish a transport core network if ten small packet-transmission nodes, such as an IP node, are connected to each PTN.

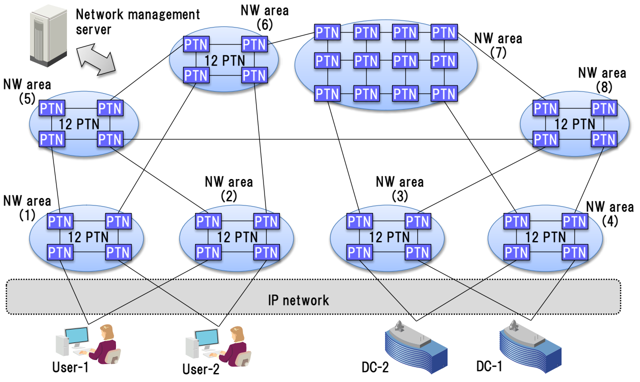

Figure 12. Evaluation system

TABLE III. Specifications of Server

#	Item	Specifications
1	CPU	1.8 GHz, 4 cores
2	Memory	16 GB
3	Storage	600 GB

B. Evaluation condition

The time taken to calculate PWs by using two routes between the users and the application servers in the DCs was evaluated. As an evaluation condition, multiple LSPs between the users and the application servers were established. Each LSP includes 10 PWs (since it usually includes multiple PWs). The evaluations were executed according to the patterns listed in Table IV. Specifically, the time required to calculate current paths and recovery paths for 255 area-based network-failure patterns was evaluated by changing the number of PWs (namely, 50+50, 250+250, and 500+500) requested by the two users. The time required to distribute all calculated recovery-path configurations and recovery IDs was also evaluated.

TABLE IV. EVALUATION ITEMS

#	Item	Specifications
1	Current-path calculation time	Time taken to calculate 50+50, 250+250, and 500+500 PWs
2	Recovery-path calculation time	Time taken to calculate recovery 50+50, 250+250, and 500+500 PWs for 255 possible area-failure patterns
3	Distribution time	Time taken to distribute all calculated recovery PWs and LSPs for 255 possible area-failure patterns
4	Recovery-ID distribution time	Time taken to distribute a recovery ID after detecting a first area failure

C. *Evaluation results*

The prototype system was evaluated according to the conditions described by the previous section.

1) Current-path calculation time

The times taken to calculate current PWs requested by the two users are plotted in Figure 13. The evaluation condition is that 10 PWs are included in one LSP. As shown in the figure, the times taken to calculate 100 (50+50) current PWs, 500 (250+250) current PWs, and 1000 (500+500) current PWs were about 64, 326, and 710 milliseconds, respectively.

2) Recovery-path calculation time

The times taken to calculate all recovery PWs for 255 possible area-based network-failure patterns by using one route are plotted in Figure 14. The evaluation condition is that 10 PWs are included in one LSP. As shown in the figure, the time taken to calculate all recovery PWs for 255 area-based network-failure patterns and 100 (50+50) current PWs, 500 (250+250) current PWs, and 1000 (500+500) current PWs are about 5.0, 31.2, and 91.1 seconds, respectively.

3) Distribution time for recovery paths

The times taken to distribute all configurations of calculated recovery PWs to all PTNs are plotted in Figure 15. The evaluation condition is that 10 PWs are included in one LSP. As shown in the figure, the times taken to distribute all configurations of recovery PWs for 255 area-based network-failure patterns and the 100 (50+50) current PWs, 500 (250+250) current PWs, and 1000 (500+500) current PWs are about 239, 315, and 427 milliseconds, respectively.

4) Recovery time from first area-based-network failure

The times taken to distribute the recovery ID to related PTNs and recover from the first area-based network failure for 100 (50+50) current PWs, 500 (250+250) current PWs, and 1000 (500+500) current PWs are plotted in Figure 16. The evaluation condition is that 10 PWs are included in one LSP. Two area-based network-failure patterns, namely, failures of network areas (1), (4), and (6) and failures of network areas (2), (5) and (8), were evaluated since they include other types of one or two area-based network-failure patterns. As shown in the figure, in the case of failures of network areas (1), (4), and (6), the times taken to recover from the first failure for 100 (50+50) current PWs, 500 (250+250) current PWs, and 1000 (500+500) current PWs are about 596, 1625, and 3332 milliseconds, respectively. On the other hand, in the case of the failures of network areas (2), (5), and (8), the times taken to recover for 100 (50+50) current PWs, 500 (250+250) current PWs, and 1000 (500+500) current PWs are about 649, 1913, and 3513 milliseconds, respectively. As shown in the figure, even if 1000 PWs are setup, the system could recover from the three area failures within four seconds. However, the time taken to recover from the first detected area failure depends on the number of setup PWs. The reason for that dependence seems to be that it takes some time to detect another area failure because the recovery procedures are begun after the first failure is detected. Consequently, the more area-based network failures occur, the longer the time taken to recover from them.

5) Time for recovery-path configuration when number of PWs is changed

The times taken to distribute the recovery ID to related PTNs and setup recovery paths (such as PWs and LSPs) after detecting the first area-based network failure (“configuration time” hereafter) were evaluated (see Figure 17). The evaluation condition is that 10 PWs are included in one LSP. As shown in the figure, the configuration time for 100 recovery PWs, 500 recovery PWs, and 1000 recovery PWs requested by a user are about 70, 94, and 74 milliseconds, respectively. In any case, the configuration time is under 100 milliseconds, even in the case of 1000 PWs. This evaluated time is regarded as the “pure” configuration time for recovery (namely, the time needed for distributing recovery ID and setting up recovery paths, excluding the time taken to detect network failures). In the figure, for a comparison with the proposed method, the time taken to calculate 1000 PWs by a conventional restoration method is also depicted. With the conventional method, calculation of 1000 PWs takes 769 milliseconds. As shown the figure, compared to the proposed method, the conventional method takes much more time to prepare recovery paths.

6) Time for configuring recovery paths when number of LSPs is changed

In the previous experimental evaluation, configuration times for various numbers of setup PWs were evaluated. In this evaluation, the configuration times for various numbers of setup LSPs were evaluated (see Figure 18). Three cases were evaluated. In the first case, namely, 1000 LSPs, a PW is accommodated in each LSP. In the second case, namely, 100 LSPs, 10 PWs are accommodated in a LSP. In the third case, namely, 10 LSPs, 100 PWs are accommodated in a LSP. In all cases, 1000 PWs are setup as recovery paths. As shown in the figure, the configuration times for 1000 current LSPs, 100 current LSPs, and 10 current LSPs requested by a user are about 90, 64, and 64 milliseconds, respectively. In all cases, the configuration time is under 100 milliseconds, even in the case of 1000 PWs.

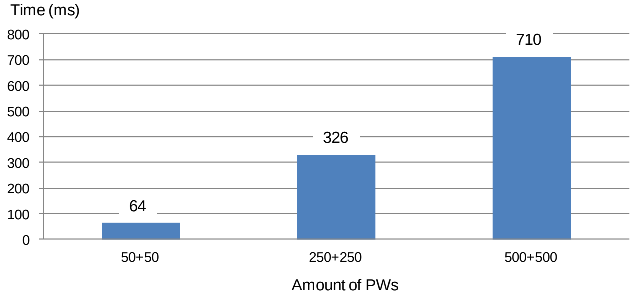

Figure 13. Time for calculating current paths

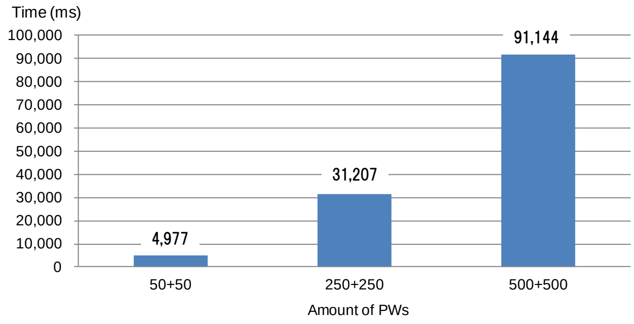

Figure 14. Time for calculating recovery paths

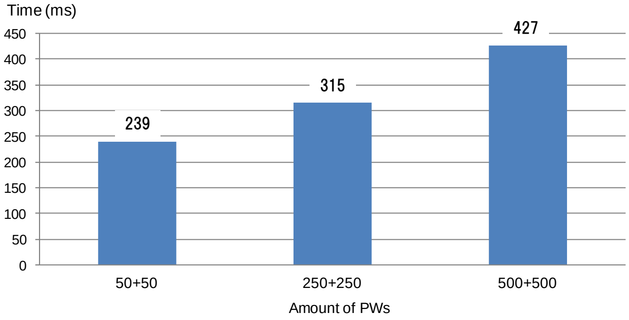

Figure 15. Time for distributing recovery paths

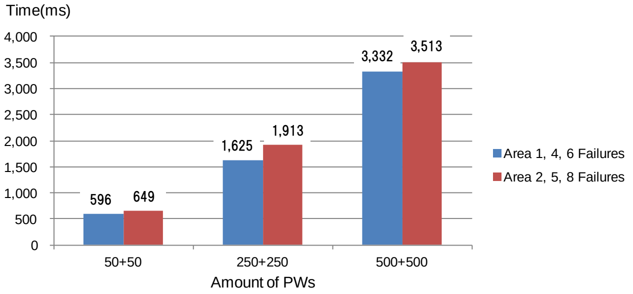

Figure 16. Time for recovery from first area failure

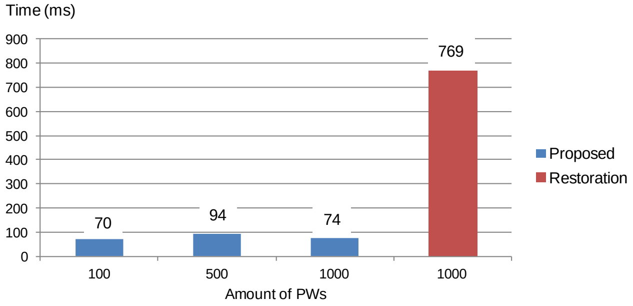

Figure 17. Time for configuring recovery paths when number of PWs is varied

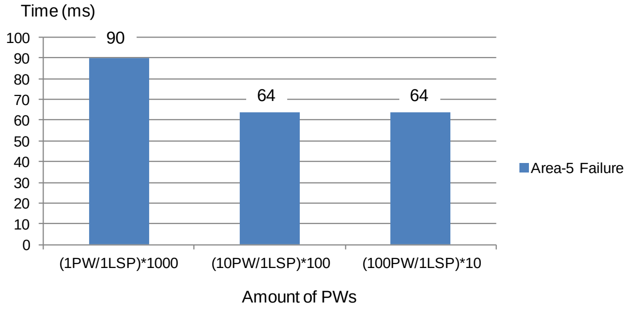

Figure 18. Time for configuring recovery paths when number of LSPs is varied

D. Comparison of proposed system and conventional system

A restoration scheme [2] is basically used when catastrophic network failures occur. In other words, a large number of setup paths are recalculated after network failures are found. According to Figure 13, it takes about 710 milliseconds to calculate paths for 1000 (500+500) PWs. If 1,000,000 PWs exist, it may take 11 minutes and 50 seconds to calculate all the paths. That is, over 10 minutes are needed to calculate recovery paths for the 1,000,000 PWs setup after the network failures were found. On the other hand, in the case of the proposed system, information needed for recovering all the setup paths is distributed to all the network nodes (such as PTNs). According to Figures 17 and 18, the recovery-path configuration time after the network failures are found is less than 100 milliseconds in all cases, since the configuration time is basically independent of the number of setup PWs. Therefore, even if 1,000,000 PWs exist, the proposed system can start recovery within 100 milliseconds after network failures are found.

On the condition that the configuration time depends on the size of the recovery-path database, it takes about 1 minute and 40 seconds to configure 1,000,000 PWs, since 1000 PWs are configured within 100 milliseconds by the proposed system under the conditions specified in Figure 12. Even if the database is very large, recovery-path configurations for all setup paths are selected only once.

Namely, it is enough to select one backup operation plane specified by the recovery ID from the prepared multiple-backup operation planes. It is therefore supposed that the relation between recovery-path configuration time and number of setup paths is almost linear. If 100,000 PWs exist, the proposed system can start to recover paths within 10 seconds. On the other hand, a conventional system based on a restoration scheme takes 1 minute and 12 seconds.

With regard to cost, compared to conventional systems (which use a restoration scheme), the proposed system needs more memory (storage) capacity to keep the recovery paths. As a rough estimation, if the size of the path-configuration data is 1 Kbyte and one-million paths and 1000 backup operation planes exist, each node in the proposed system needs 1 Tbyte of storage (depending on the established paths). However, memory and/or storage costs have been gradually decreasing, so the proposed system is promising for application in the near future.

VII. Conclusion

A "network-disaster recovery system" using area-based network management is proposed. As for this system, a whole network is separated into multiple areas. Each area is composed of multiple network nodes, such as MPLS-TP nodes. The system is managed by a network-management server that monitors the condition of every network node and manages the network by detecting area-based failures. Before starting network operations, it calculates recovery paths for every possible area-based failure and distributes them with a recovery ID for each area-failure pattern. The network nodes receive and store the recovery-path configuration and recovery ID. The network-management server detects the network-area failures during network operations and determines a pattern of area failures. Specifically, it determines the numbers and positions of area failures. After determining the pattern of area failures, the network-management server selects an appropriate recovery ID for that pattern and distributes the ID to recovery-related network nodes. The network nodes receive the recovery ID and start data transmission based on the path configuration specified by the distributed ID. After these procedures are completed, the area failures are swiftly recovered.

A prototype system, composed of a network-management server and 96 simulated packet-transport nodes, with a graphical viewer was implemented, and its performance was evaluated. According to the results of the evaluation, all recovery-path configurations for 1000 PWs, namely, transmitting the recovery ID to the related network nodes and using a recovery-path database specified by the ID, are done within 100 milliseconds after network-area failures are detected. On the condition that the configuration time depends on the size of the recovery-path database, the proposed system takes about one minute and 40 seconds in the case of 1,000,000 PWs. On the other hand, it takes a conventional restoration scheme over 10 minutes to calculate recovery paths under the same evaluation conditions used for the proposed system.

As for the prototype system, the whole network is divided into eight areas as one example of dividing the whole network into multiple area networks. However, the scalability of this approach is an issue. For example, an extended recovery scheme is needed when only one link or node failure occurs, since the proposed system is useful for a large-scale network and multiple failures. The system should be useful for both small failures and large failures. In addition, a consistency of database between a node and a management-server is a future issue. Besides, a recovery procedure is needed in case of a failure of a management server. Therefore, the recovery scheme will be further developed.

Acknowledgments

Part of this research was included in Research Project O_3 (Open, Organic, Optima) and was supported by the MIC (Japanese Ministry of Internal Affairs and Communications) program, "Research and Development on Virtualized Network Integration Technology".

References

[1] T. Suzuki et al., "A network-disaster recovery system using area-based nework management," The Third International Conference on Communications, Computation, Networks and Technologies (INNOV 2014), pp. 8-15, Oct. 2014.

[2] E. Mannie and D. Papadimitriou, "Recovery (protection and restoration) terminology for generalized multi-protocol label switching (GMPLS)," RFC 4427, Mar. 2006.

[3] International Telecommunication Union - Telecommunication Standardization Sector (ITU-T)

http://www.itu.int/en/ITU-T/Pages/default.aspx [retrieved: June, 2015].

[4] The Internet Engineering Task Force (IETF),

http://www.ietf.org/ [retrieved: June, 2015].

[5] B. Niven-Jenkins, D. Brungard, M. Betts, N. Sprecher, and S. Ueno, "Requirements of an MPLS transport profile," RFC 5654, Sept. 2009.

[6] M. Bocci, S. Bryant, D. Frost, L. Levrau, and L. Berger, "A framework for MPLS in transport networks," RFC 5921, July 2010.

[7] T. Busi and D. Allan, "Operations, administration, and maintenance framework for MPLS-based transport neworks," RFC 6371, Sept. 2011.

[8] N. Sprecher and A. Farrel, "MPLS transport profile (MPLS-TP) survivability framework," RFC 6372, Sept. 2011.

[9] P. Pan, G. Swallow, and A. Atlas, " Fast reroute extensions to RSVP-TE for LSP tunnels," RFC 4090, May 2005.

[10] M. Pickavet, P. Demeester, and D. Colle, "Recovery in multilayer optical networks," Journal of Lightwave Technology, Vol. 24, no. 1, pp. 122-134, Jan. 2006.

[11] J. Zhang, J. Zhou, J. Ren, and B. Wang, “A LDP fast protection switching scheme for concurrent multiple failures in MPLS network,” 2009 MINES '09. International Conference on Multimedia Information Networking and Security, pp. 259-262, Nov. 2009.

[12] Z. Jia and G. Yunfei, “Multiple mode protection switching failure recovery mechanism under MPLS network,” 2010 Second International Conference on Modeling, Simulation and Visualization Methods (WMSVM), pp. 289-292, May 2010.

[13] G. Kuperman and E. Modiano, “Network protection with guaranteed recovery times using recovery domains,” INFOCOM, 2013 Proceedings IEEE, pp. 692-700, April 2013.

[14] J. Rack, “Fast service recovery under shared protection in WDM networks,” Journal of Lightwave Technology, Vol. 30, no. 1, pp. 84-95, Jan. 2012.

[15] M. Lucci, A. Valenti, F. Matera, and D. Del Buono, “Investigation on fast MPLS restoration technique for a GbE wide area transport network: A disaster recovery case,” 12th International Conference on Transparent Optical Networks (ICTON), Tu.C3.4, June 2010.

[16] D. Sheela, M. Smitha Krishnan, and C. Chellamuthu, “Combined link weight based restration strategy in optical networks,” 2012 International Conference on Computing, Electronics and Electrical Technologies (ICCEET), pp. 687-690, Mar. 2012.

[17] T. S. Pham. J. Lattmann, J. Lutton, L. Valeyre, J. Carlier, and D. Nace, “A restoration scheme for virtual networks using switches,” 2012 4th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT), pp. 800-805, Oct. 2012.

[18] X. Wang, X. Jiang, C. Nguyen, X. Zhang, and S. Lu, “Fast connection recovery against region failures with landmark-based source routing,” 2013 9th International Conference on the Design of Reliable Communication Networks (DRCN), pp. 11-19, Mar. 2013.

[19] E. Rosen, A. Viswanathan, and R. Callon, “Multiprotocol lable switching architecture,” RFC 3031, Jan. 2001.

[20] S. Bryant and P. Pate, “Pseudo wire emulation edge-to-edge (PWE3) architecture,” RFC 3985, Mar. 2005.