GMPLS WITH INTERLAYER CONTROL FOR SESSION-UNINTERRUPTED DISASTER RECOVERY ACROSS DISTRIBUTED DATA CENTERS

Tetsuo IMAI[1], Soichiro ARAKI[2], Tomoyoshi SUGAWARA[3], Norihito FUJITA[4], and Yoshihiko SUEMURA[5]

[1-5] *System Platforms Research Laboratories, NEC Corporation,*
1753, Shimonumabe, Nakahara-ku, Kawasaki, Kanagawa 211-8666, Japan,
{t-imai@cl[1], s-araki@cj[2], tom-sugawara@ap[3], n-fujita@bk[4], y-suemura@bp[5]}.jp.nec.com

Abstract: We propose a session-uninterrupted disaster recovery system using a novel session migration technique as a GMPLS application. Existing disaster recovery systems have a problem of a service interruption. The session migration based on an interlayer control of GMPLS, VLAN change-over, and process migration, maintains continuous TCP service between a user and a virtualized server, even when the service migrates from a primary data center to a backup one. We developed a prototype system and showed that BoD (bandwidth on demand) by GMPLS improved the recovery time from 80.10 sec to 9.85 sec, during transmitting a process data of 40MByte.

1. INTRODUCTION

It is important for mission-critical businesses to continue even in the event of a disaster that may bring a whole data center to a halt. In this paper, we propose a novel session-migration technique for the disaster recovery system in data centers using the BoD (bandwidth on demand) service provided by GMPLS (generalized multi-protocol label switching) [1].

Existing disaster recovery systems are classified into following 3 types.

- Remote data backup: Make a copy of data at backup site by users(low-class)
- Data replication: Make a copy of data automatically(middle-class)

- Clustering: The same system is set up at backup site, and make a copy of data automatically(high-class)

Our session-migration technique is classified as the best-class. It can rapidly change a service-providing server to an alternative backup one and change a user's network to another one at the same time, enabling users to keep their own TCP sessions. It enables data center managers to provide users with continuous service. We developed a session migration controller to control change-overs in the service-providing server and users' networks, and dynamic reservation of a bandwidth by GMPLS. And we demonstrated the effectiveness of our session-migration system.

2. SESSION MIGRATION

2.1 Problems with existing method of disaster recovery

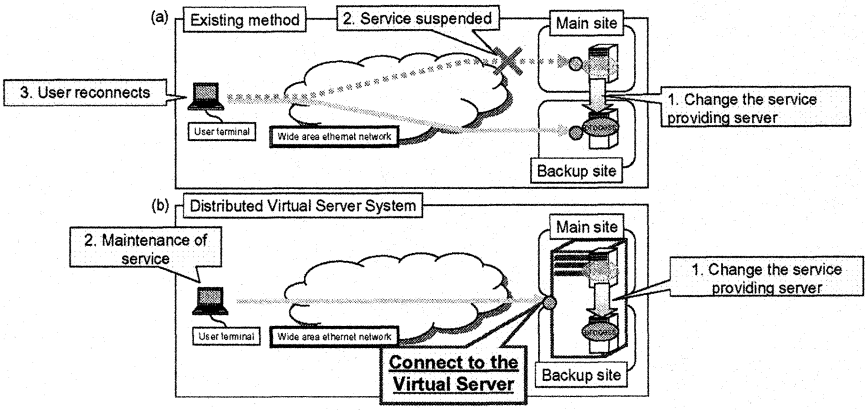


Figure 1. (a) existing disaster recovery system, (b) our distributed virtual server system.

It is essential that mission-critical businesses have continuous access to services. However, because existing recovery methods mainly use systems within the same network (subnet), they can only change the server running a service from the main site to a backup site. Therefore, there is a following problem. When a disaster occurs, the server running the process to be recovered is moved from the server at the main site (primary server) to one at a backup site (backup server) which is widely separated. To continue using the service, users have to carry out troublesome task for recovery, including closing the TCP (transmission control

protocol) session to the primary server, and re-connecting to the backup server(Fig. 1 (a)). The problem is that service to users is suspended until the recovery procedures are completed. Our proposed disaster recovery system intends not to affect users' access to services. Our session-migration technique enables TCP services between users and a virtualized server to continue, even when the service migrates from a primary server to a backup one(Fig. 1 (b)).

2.2 Session migration and current performance problems

To provide continuous service even in the event of a disaster, we migrate service-providing processes from the primary server to a backup one, which is widely separated, with maintaining the TCP session between the user and the process. A "process migration"[3] is used to swap the active server to another server by migrating a memory image of the process from one server to the other.

There are several problems in maintaining users' TCP sessions during process migration. First, we have to move the server's IP address between the servers at the same time as the process migration occurs, but sharing the same IP address in the same wide-area network creates difficulties. Secondly, when we carry out process migration between widely separated servers, we have to minimize transmitting time. Session migration is triggered by the occurrence of some sort of disasters, for example, a major electrical power failure or a fire alarm at a data center. Therefore, session migration has to begin with these warnings, and be completed before the disaster eventuates. In addition, long downtime of the process may cause a performance degradation or a session disconnection.

To solve the first problem, we distribute the same IP address to different VLANs, and we change-over the "user VLAN" (the VLAN that the user belongs to) in synchronization with process migration in a session-migration sequence. This enables users to connect quickly with the new network without affecting their use of the service. It can maintain the TCP session without interruption.

To solve the second problem, we rapidly reserve a wide bandwidth for process migration in advance by GMPLS technique.

There has been considerable work recently on new services such as BoD for optical networks using GMPLS techniques. GMPLS will enable service providers to quickly deliver various types of paths to customers[2]. In our session-migration technique, we use GMPLS to reserve a bandwidth for the process migration, which requires rapid transmission of a huge amount of data to minimize the length of time that services provided by a data center are suspended.

And it is important to coordinate these three techniques of different layers, the change-over of user VLAN, the process migration, and the GMPLS. In our session migration, an MCS (migration-control server) controls each of them.

A detailed description of our session-migration technique follows.

2.3 Overall system architecture

Figure 2 shows the architecture of the session-migration system. "Server 1" is at the main site, and "Server 2" is at the backup site; a PMC (process migration controller) is implemented in both of them. There are routers at the edge of each site. Users and sites are connected by a wide-area Ethernet network consisting of L2SWs (layer 2 switches that support VLANs), and an underlying L1NW (layer 1 network) consisting of L1SWs (layer 1 switches) controlled by respective CP (control plane)-devices. In a wide-area Ethernet network, user sites and the main site are connected by a VLAN 1, and user sites and the backup site are connected by a VLAN 2. In the dedicated control network, an MCS is connected to a PMC, which controls process migration, a NAGW (network access gateway), which controls changes in the user VLAN, a CP-device, which directs the GMPLS to reserve a bandwidth in the L1NW, and the L2SW, which supports the VLANs.

Ordinarily, a user connects to a virtual IP address, 192 168.2.1, at Server 1 in Main site via a NAGW to utilize a service. In the following description, we use "192.168.2.1" as the virtual IP address, but it's tentative.

The components in Fig. 2 are described below.

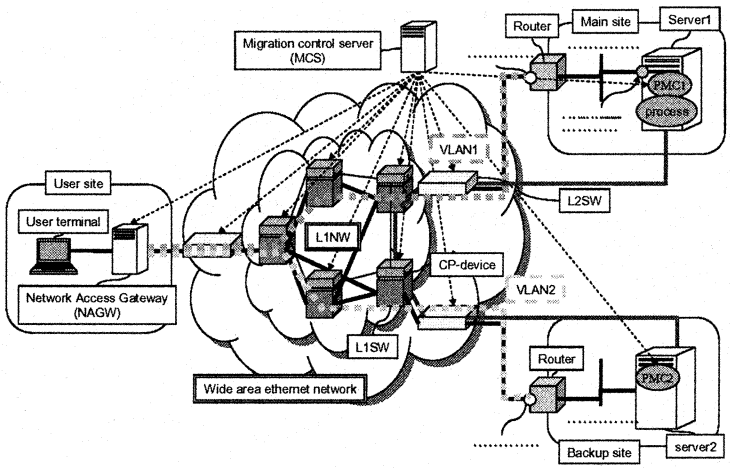


Figure 2. System Architecture of the session migration.

The MCS manages the session-migration sequence, and directs the behavior of other elements of the session-migration process through remote commands using a Telnet or other methods. The PMC 1 and PMC 2 control process migration when a

trigger is given by the MCS. The process migration technique transports a running task (process) from one server to another, including the state of communication and virtual IP address, and enables the process to keep running on the other server. The NAGW manages the rapid switch-over of server required to be connected by changing the user VLAN; i.e., the NAGW changes the user VLAN from VLAN 1 to VLAN 2, as shown in Fig. 2, when a trigger is given by the MCS. The CP-devices control the L1SWs via the GMPLS. They carry out reservation/deletion of the bandwidth between Server 1 and Server 2 when a trigger is given by the MCS, and can dynamically establish a L1-VPN (layer 1-virtual private network). Both of the L2SWs support VLANs and can dynamically establish a L2-VPN between Server 1 and Server 2 when a trigger is given by the MCS. The routers at the main site and the one at the backup site are in subnets of the same address architecture. Therefore, there is no need to change the routing table for the routers when a virtual IP address is moved from Server 1 to Server 2.

2.4 Detailed sequence of session migration

The flow of the session-migration sequence is shown in Table 1. Ordinarily, a user connects to the virtual IP address 192.168.2.1 of Server 1, and utilizes a service by communicating with a process on Server 1.

When the alarm sounds for a disaster, the data center manager, or automatic equipment, commands the MCS to begin session migration (STEP 1). The MCS commands the CP-devices and L2SWs to configure a L1/L2-VPN (STEP 2, 3). It set up the dedicated and wideband VPN for process migration.

When the VPNs are set up, the MCS starts the sequence for process migration. The MCS commands the PMC 1 to freeze the process (STEP 4), and communication between the user and Server 1 is suspended. Next, the MCS commands PMC1 and PMC 2 to migrate the process (STEP 5), and to move the virtual IP address 192.168.2.1 from Server 1 to Server 2 (STEP 6). This causes temporary disappearance of the IP address 192.168.2.1 to the user because the user is connected to VLAN 1 at this point. At STEP 7, the MCS commands the NAGW to change the user VLAN from VLAN 1 to VLAN 2, and the user rediscovers the IP address 192.168.2.1. Throughout this sequence, the user simply loses the IP address and rediscovers it, and is unaware of the change in the VLAN to which the user is connected. Then the MCS commands the PMCs to resume the process(STEP 8); the user can then resume communicating with the server (Server 2) and using the service.

Finally, the MCS commands the CP-devices and L2SWs to delete L1/L2-VPN (STEP 9, 10), and reports the completion to the data center manager (STEP 11).

In the aspect of effect of session migration on users, the following issues can potentially affect users during service migration.

Processing period from STEP 4 to STEP 8: User communication to the server is suspended during this time. However, depending on the type of service, if this period is very brief, it does not affect service utilization in general.

Processing time required for STEP 7: The user loses the virtual IP address 192.168.2.1 of the servers at this point. Again, depending on the type of service, if this time is very brief, the user's application will not detect packet convergence, and will not stop the use of the service.

If both these processing periods are sufficiently brief, the user will be unaware of the session migration and will experience continuous service.

Table 1.Sequence for session migration.

STEP 1:	The manager commands the MCS to start session migration.	Pre process
STEP 2:	The CP-devices reserve a bandwidth for process migration.	
STEP 3:	The L2SWs configure the VLAN for process migration.	
STEP 4:	The PMC freezes the process to be migrated.	Main process
STEP 5:	The PMC carries out process migration.	
STEP 6:	The PMC changes over the virtual IP address.	
STEP 7:	The NAGW changes the user VLAN from VLAN 1 to VLAN 2.	
STEP 8:	The PMC resumes the process.	
STEP 9:	The CP-devices deletes the bandwidth.	Post process
STEP 10:	The L2SW delete the VLAN.	
STEP 11:	The MCS reports the completion of session migration.	

3. PROTOTYPE SESSION MIGRATION SYSTEM AND ITS PERFORMANCE EVALUATION

We built a prototype system (Fig. 3) to evaluate the session migration technique described above. The MCS, L2/L3SW, and CP devices were constructed on a rack on the right of the photograph; on the left, there is a rack of SDH nodes (SpectralWave U-node, NEC) as the L1SW controlled by the CP devices. The NAGW is installed in the PC at the right. The features of these components are listed in Tables 3 and 4. Server 1 and Server 2 have a Linux OS, each PMC is built as a daemon on Linux. The process migrated in the demonstration was a video data transmission process, which uses 40 MB of memory; the user receives streaming data via a TCP at a bit rate of 2 Mbps. Server 1 and Server 2 are connected by a gigabit Ethernet to the individual SDH nodes, which are connected to each other by a 2.4-Gbps link. During the session migration sequence, the GMPLS commands the SDH nodes to reserve a bandwidth of 1 Gbps. It then establishes a bandwidth of 1 Gbps between servers.

We carried out session migration using this architecture and found that the procedure did not have any effects on users browsing a streaming video.

Below, we discuss the effects of session migration on users' service utilization based on the discussion in section 2.4.

For comparison, we carried out a demonstration of session migration without GMPLS. In the case without GMPLS for reservation of bandwidth, servers carry out the process migration through a public network. We emulate that by connecting the L2/L3SWs with a 10-Mbps link directly; the bandwidth between the servers was then 10-Mbps. Table 2 shows the processing time required for each step of the session-migration process, with and without GMPLS. Note that STEP 2, 3, 9, 10 aren't necessary in the case without GMPLS.

As shown in Table 2, with GMPLS, the processing time from STEP 4 to STEP 8 (the period for which user communication with the server is suspended) was about 2.7 sec. We used Real One Player (Vers. 2.0) as the video browser, and it has a 30-sec buffer by default, which is sufficient to cover the suspension period.

Without GMPLS, the period was about 79.6 sec. long, which was too long to be covered by the buffer. The video stream was therefore suspended briefly.

The processing time required for STEP 7, when the user loses the virtual IP address of the server, is about 0.1 sec. in each demonstration. This demonstration showed that the video browser did not suspend service and the TCP congestion controller did not decrease its throughput.

Table 2. Processing time required for each step.

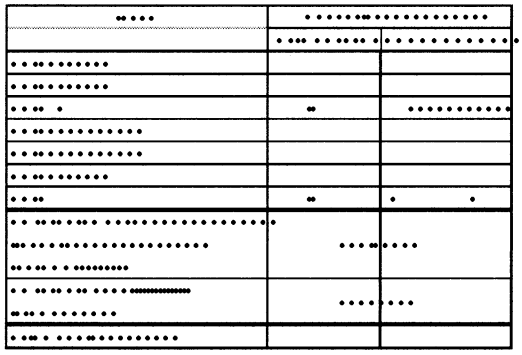


Table 3. Features of components

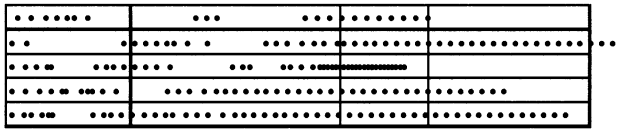
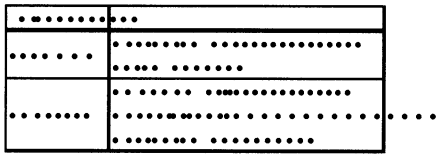


Table 4. Features of switches



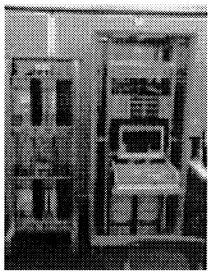


Figure 3. Demonstration system.

4. CONCLUSION

We proposed a session-uninterrupted disaster recovery system as a GMPLS application. Using the interlayer controller, our session migration technique coordinates three methods: BoD by GMPLS, process migration between servers, and a change in the user VLAN, and avoids interruption of users' accesses to the services provided by the servers. In a demonstration of the system, we showed that the technique worked well and that GMPLS improved the recovery time from 80.10sec to 9.85 sec during transmitting a process data of 40MByte.

ACKNOWLEDGMENTS

This work includes a part of the result of Key Technology Reseach Promotion Project (the research project on large-scale high reliability server) supported by NEDO.

REFERENCES

[1] E. Mannie, "Generalized Multi-Protocol Label Switching Architecture," IETF Internet Draft, draft-ietf-ccamp-gmpls-architecture-07.txt.

[2] H. Ishimatsu, S. Tanaka, M. Akashi, T. Hashimoto, E. Yamaguchi, Y. Oyama, H. Nakano, A. Inomata, M. Murakami, Y. Ashikaya, S. Ryu, "Prototype Demonstration of On-Demand/Scheduled Wavelength Path Service," OFC2003, ThR2, 2003.

[3] Dejan S. Milojicic, F. Douglis, Y. Paindaveine, R. Wheeler, S. Zhou, "Process Migration," In ACM Computing Surveys, Volume 32, Issue 3, September 2000.