OSS Distributed System Disaster Countermeasure

Technology Reports

Improving Operations System Reliability in Large-scale Disasters

The Great East Japan Earthquake caused large-scale damage over a wide area severely impacting a variety of social infrastructures. For Japan, such an unprecedented disaster underscored the importance of disaster countermeasures in society. The social mission of NTT DOCOMO as a telecommunication carrier is to provide stable network quality even at the time of a large-scale disaster. In particular, its OSS that supports the operation and maintenance of the network must be able to correctly assess network conditions during a disaster and provide appropriate operation and maintenance functions. This article describes the disaster countermeasures provided by NTT DOCOMO's OSS.

Core Network Development Department

Kosuke Kagawa
Yuya Kuno
Hironao Tamura
Hisashi Takada
Masanori Furutani
Nobuya Minakata

1. Introduction

The Great East Japan Earthquake that struck in March 2011 caused a massive amount of damage over a wide area. The communications network was no exemption: many Network Element (NE)[*1] blocks failed and many transmission paths suffered major damage including severed cables especially in the stricken area. Fortunately, NTT DOCOMO's Operation Support System (OSS)[*2] itself was not directly damaged this time, but the disaster brought to the forefront new requirements for providing stable network operations and appropriate maintenance functions at the time of a large-scale disaster.

In this article, we explain OSS requirements for dealing with a large-scale disaster and describe the methods for achieving them.

2. New Requirements from Disaster Experiences

2.1 Problems

1) NE

The Great East Japan Earthquake caused temporary service interruptions in approximately 5,000 base stations especially in the stricken area while also causing many long-term power outages and equipment failures throughout eastern Japan.

At the same time, traffic increased by 50 – 60 times that of normal times as calls were made within the stricken area to exchange information and as people from the outside called into the area to check on the well-being of loved ones, friends, colleagues, etc. This traffic generated an abnormal load on the NEs making up the network and created conditions in which a congestion control function had to be activated in each NE throughout the country to prevent

*1 **NE**: A functional block that achieves a necessary function in the provision of telecommunication services; specifically, a unit of telecommunication equipment such as a switch, transmitter or radio station.

congestion[*3] throughout the network.

2) OSS

Network-monitoring conditions at the time of the disaster are shown in **Figure 1**.

Under the network conditions described above, the OSS fell into an abnormal state in which a large volume of alarms continued to be received for a relatively long period of time from multiple NEs. At this time, the system experienced congestion and a lack of resources as well as some data loss and processing delays. This state of affairs made it difficult to quickly assess failure conditions in the network.

In addition, failures and congestion-control operations occurred in many base stations generating a huge amount of information related to those failures and the state of congestion control. This information was needed if damage conditions were to be understood, but the sheer amount of data involved made it difficult to draw up an efficient plan for area restoration including a restoration priority sequence or to accurately determine load conditions across the entire network.

3) Disaster Recovery[*4]

Since the OSS needs to be able to process alarms from many NEs in real time, it has a large-scale configuration consisting of many servers.

Furthermore, to therefore prevent processing performance from deteriorating as a result of transmission-path delays in communications between

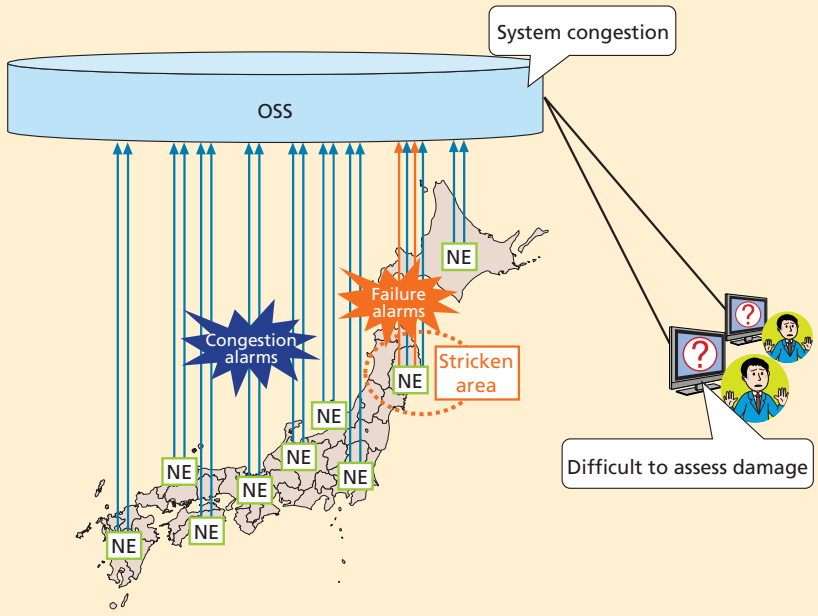


Figure 1 Network-monitoring conditions at the time of a disaster

applications, this real-time, large-scale system adopts a disaster-recovery scheme consisting of an operations system constructed at one site and a backup system constructed at another site (**Figure 2**).

In this disaster-recovery scheme, the backup system is placed in idle operation during normal times, which makes for a very low facility utilization factor. In addition, the switchover from the operations system to the backup system requires some manual switching of network NEs to the backup system in a certain sequence, which means that some time is needed before business operations can be restored.

2.2 Clarification of Requirements

The following summarizes new OSS requirements that we have established in face of the problems arising out of the recent disaster as described above.

- System congestion or crashes must not be allowed to occur even under abnormal conditions in which a large volume of alarms are being continuously received from many NEs.
- Information for determining a priority sequence for restoring failed stations must be provided.
- The state of NE congestion control must be uniformly managed and displayed so that load congestion across the entire network can be understood.
- Disaster recovery must be efficiently achieved and business operations quickly restored.

*2 **OSS**: A system for discovering failures and congestion in the mobile communications network and performing appropriate control functions or measures in response to such problems.

*3 **Congestion**: A state in which communication requests become concentrated in a short period of time thereby overwhelming the processing capability of NEs and generating communication failures.

*4 **Disaster recovery**: Repair and restoration of a system damaged by a natural disaster or other calamity. Also, preventive measures for minimizing damage.

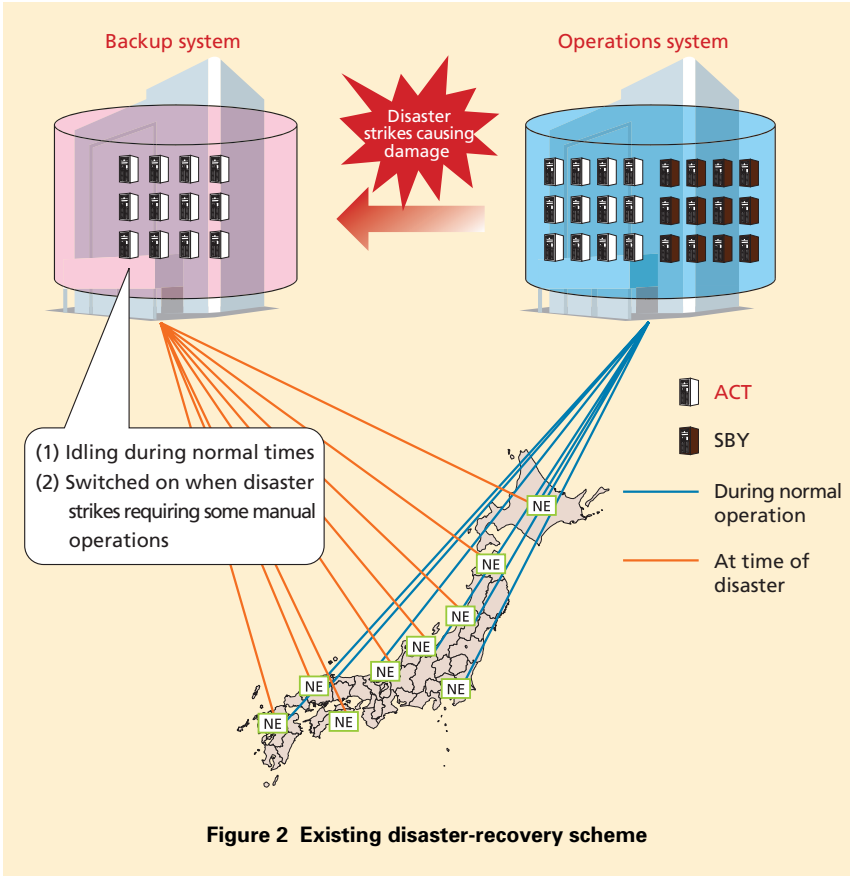


Figure 2 Existing disaster-recovery scheme

3. OSS Disaster Countermeasures

3.1 OSS Congestion Countermeasure

1) Existing Congestion Countermeasure and Problem

There are approximately 100,000 NEs monitored by NTT DOCOMO's OSS throughout Japan. The system adopts a configuration in which one server accommodates multiple NEs, which means that the failure of a single NE or the frequent issuing of alarms by a NE due to a small-scale disaster has the possibility of degrading the processing performance of the entire server. To prevent this from happening, each NE is given a congestion control function.

The existing congestion control function performs real-time monitoring of the quantity of alarms being issued from the NE in question. In the event that this quantity exceeds a certain threshold, the function will halt alarm notification from that NE. If the quantity of alarms should then drop below this threshold, it will restart alarm notification. As a result, the OSS will again be able to obtain information on the NE so that the NE and OSS will again be consistent with regard to the NE's state. In short, this form of congestion control is performed in units of NEs for a large quantity of alarm notifications generated in a short period of time.

In an abnormal situation in which a large volume of alarms are being received from many NEs over a long period of time as in the case of the Great East Japan Earthquake, the existing congestion control system will reinitiate the issuing of alarms even during NE-state gathering following the cancellation of an alarm-notification halt with the result that the congestion control function will repeatedly be invoked (**Figure 3**).

Under such conditions, this invoking of the congestion control function from many server-accommodated NEs will result in system congestion and resource depletion. To prevent this from happening, there is a need for congestion control that takes into account the possibility of a large-scale disaster.

2) New Congestion Control

To solve the above problem, we have implemented a new type of congestion control in units of servers (each of which accommodates multiple NEs) in addition to the above type of congestion control executed in units of NEs. The new congestion control system is shown in **Figure 4**. This new system incorporates the following two functions:

(1) A function for monitoring the number of alarms being issued from all NEs on the server in question to assess the state of congestion

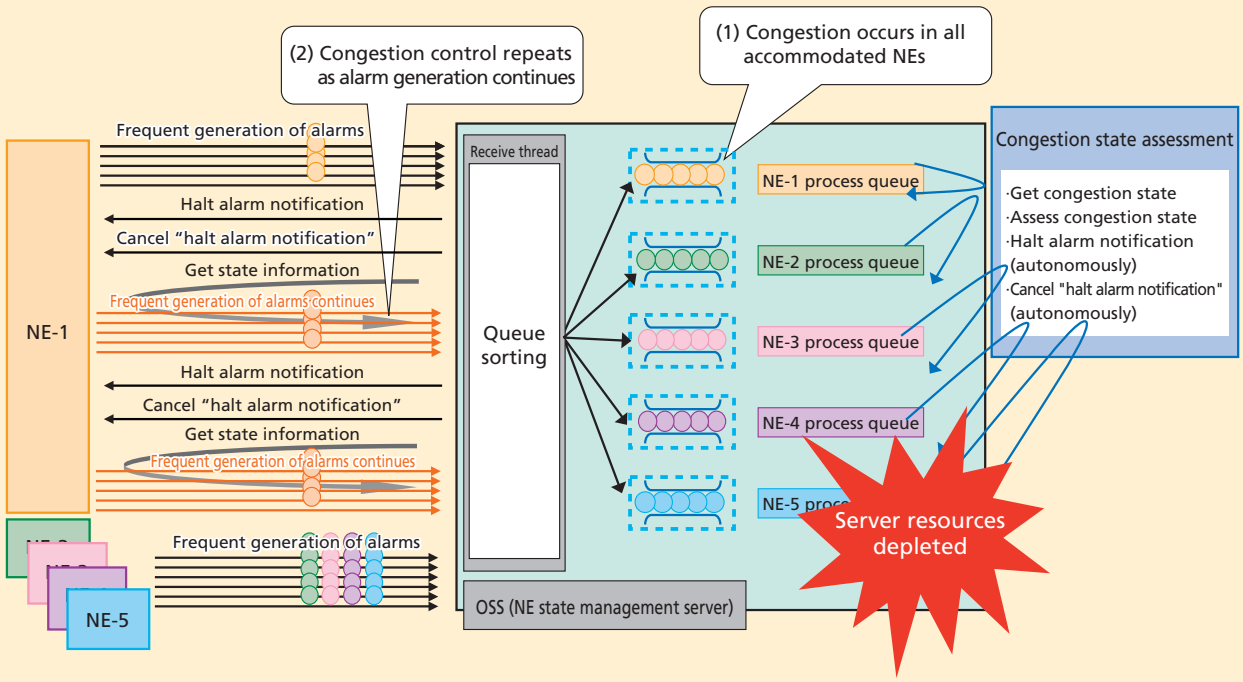


Figure 3 Problem in existing congestion control

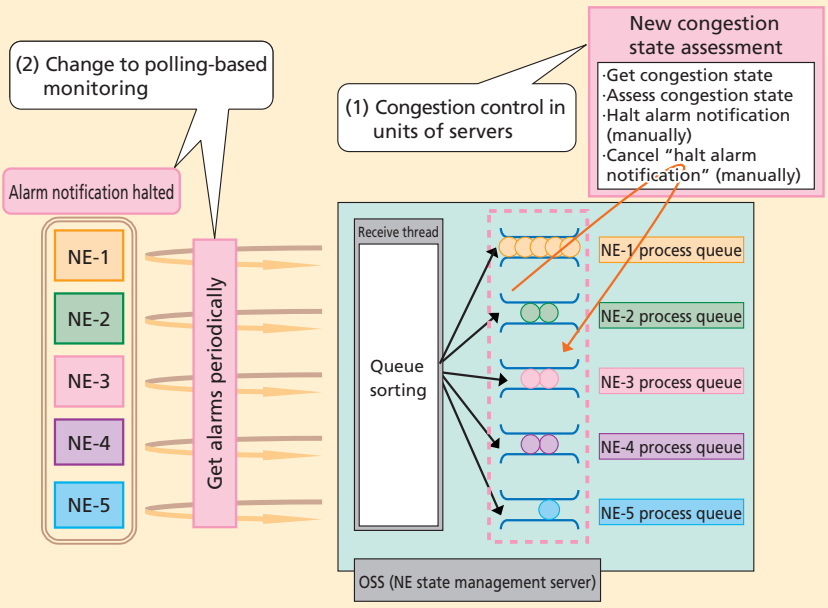


Figure 4 New congestion control system

(2) A function for switching from a monitoring system based on alarm notification from NEs to the OSS to one that performs polling[*5] from the OSS to NEs to collect information

Targeting all NEs on a server to assess congestion in this way makes it possible to prevent congestion from occurring on that server even when alarms are being issued from many NEs. Furthermore, the use of a polling system to perform monitoring at the time of a disaster will, on the one hand, degrade real-time characteristics compared to monitoring during normal times, but will, on the other hand, prevent congestion from occurring in the

*5 **Polling**: The sending of inquiries from a terminal to a server to see if data is available for transmission.

OSS even at the time of a large-scale disaster thereby enabling the state of NE failures to be reliably obtained. This will enable operators and maintenance personnel to obtain accurate information on failures in the network and to perform appropriate network control [1].

3.2 Display of Information for Determining a Restoration Priority Sequence

Up to now, the OSS has been providing information related to equipment failures. In the event of a large-scale disaster, however, the information needed for operation and maintenance will also include information that can be used for determining from which of the many failed NEs restoration should begin.

For this reason, we have made it possible to visually identify failures associated with service interruptions from the information on NE equipment failure that has traditionally been collected. Furthermore, for a failure associated with a service interruption, we have enabled the number of users affected by that NE failure to be displayed.

The number of users affected by such a service interruption is obtained by the following procedure based on traffic information collected by the OSS (**Figure 5**).

First, the number of resident users is calculated based on the number of resident profiles. That is, the number of resident users is obtained from NEs that hold such profiles. Next, considering that the number of users accommodated by each base station or sector is proportional to traffic volume, the total number of resident users is divided up among base stations in proportion to the traffic volume of each. Specifically, the number of users accommodated by each base station is calculated using the following formula:

No. of affected users: α

No. of Local Area Code (LAC)[*6] resident users: β

Traffic volume of failed NE: γ

Traffic volume of unit LAC: δ

$$\alpha = \frac{\beta \times \gamma}{\delta} \quad (1)$$

In general, the area governed by a certain base station will include an overlay of different frequencies and will also be supplemented by the area governed by a neighboring base station. In addition, the frequencies that can be received by a user terminal are limited. Additional considerations must therefore be made with respect to Equation (1) in calculating the actual number of users affected by a service interruption (**Figure 6**). In this revised calculation, the proportion of users rescued by the overlay base station and the proportion of users rescued by a neighboring base station are subtracted using the following formula:

Supported-terminal ratio: ε

Neighboring-area rescue ratio: ζ

$$\alpha = \frac{\beta \times \gamma \times (1-\varepsilon) \times (1-\zeta)}{\delta} \quad (2)$$

Since traffic data can no longer be obtained after the NE in question fails, this calculation is based on data obtained and accumulated during normal times. In actuality, the number of affected users is calculated from data obtained just prior to failure occurrence.

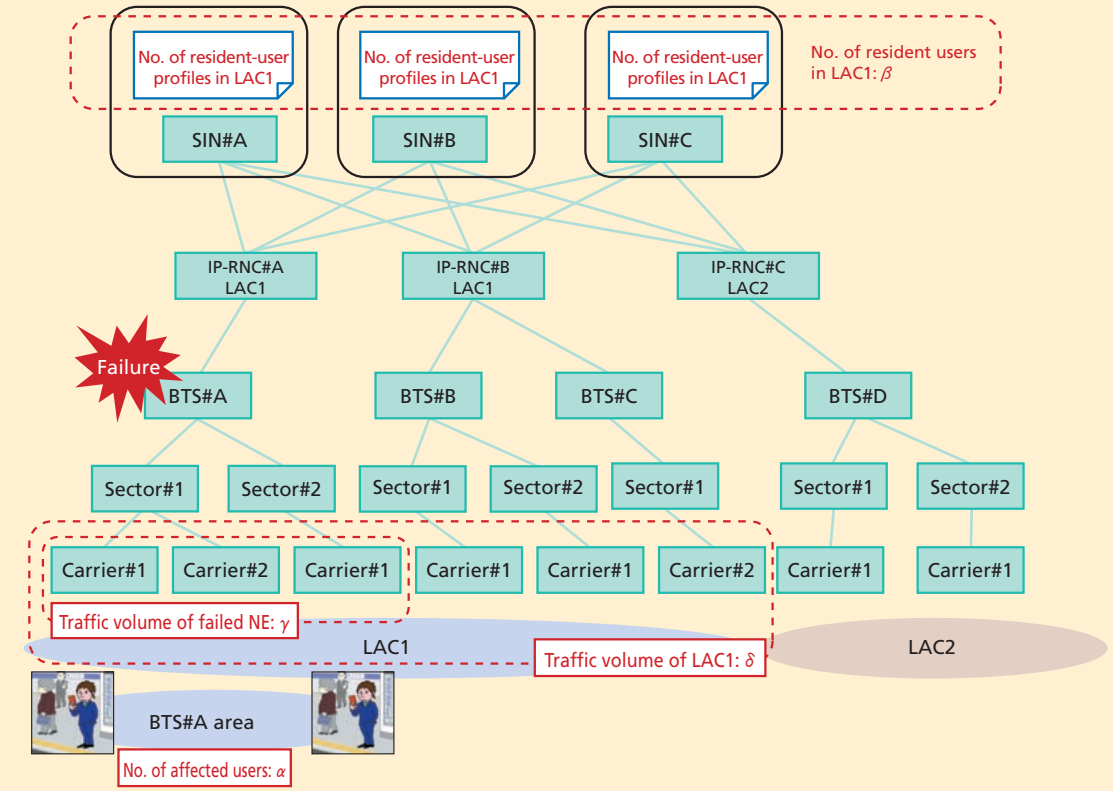
Displaying service-interruption information and number of affected users in this way simplifies the task of determining a restoration priority sequence and enables area restoration to be efficiently accomplished at the time of a large-scale disaster.

3.3 Network Control by Consolidating Congestion Control Alarms

We have implemented a visualization function to enable the state of NE congestion control throughout the network to be quickly and accurately assessed.

This function periodically obtains and analyzes the state of congestion control for each NE and provides useful information to operators and maintenance personnel. The monitoring screen provides flexible display capabilities presenting information from wide-area macro areas to micro areas as small as individual NEs. This screen also displays information on NEs as a factor in controlling congestion as well as con-

*6 **LAC**: A simultaneous paging area as a switching unit.



Signaling Interworking Node for 3G access (SIN): 3G radio-access accommodating equipment
IP-Radio Network Controller (IP-RNC): 3G radio-access control equipment
Carrier: bandwidth (In 3G, fixed at 5 MHz)

Figure 5 Method for calculating resident users

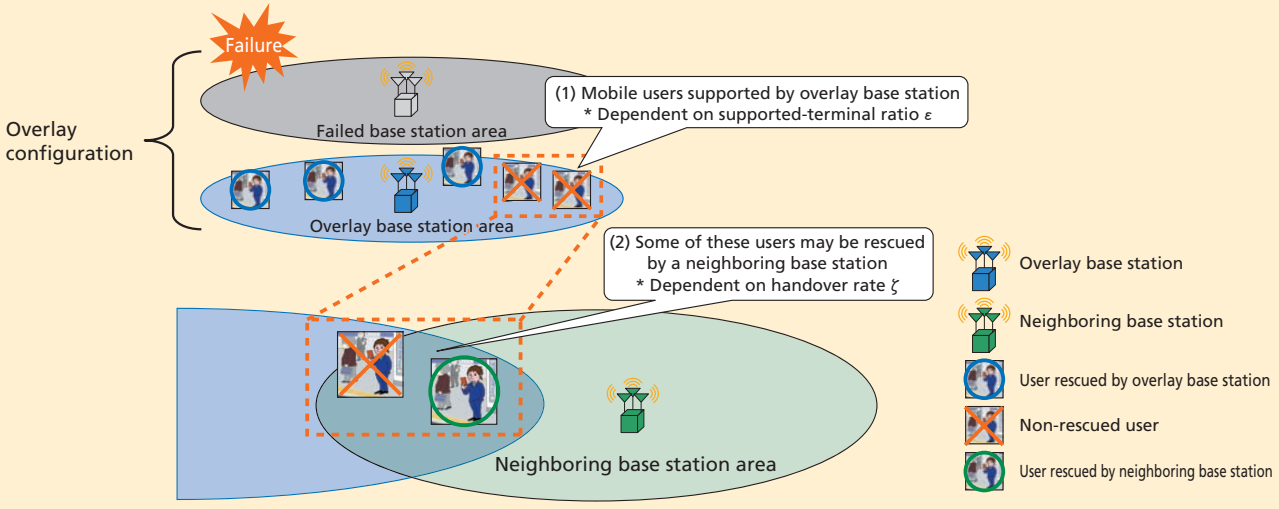


Figure 6 Calculation of users rescued by alternate base stations

gestion control type (automatic or manual) as shown in **Figure 7**.

Enabling NE congestion conditions to be visualized in this way makes for prompt and accurate network control even at the time of a large-scale disaster.

3.4 Revising the OS Disaster-recovery Scheme

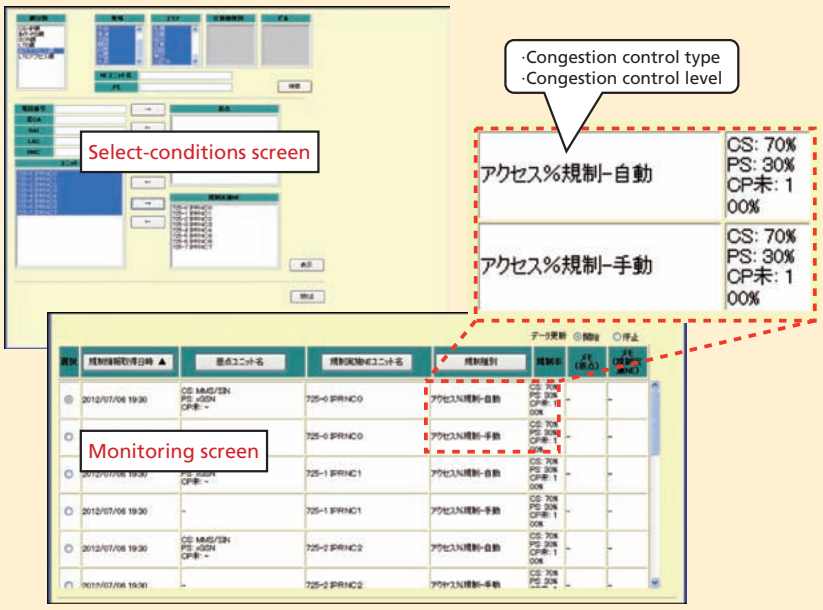
1) Refining Requirements

The requirements presented in Chapter 2 with regard to OSS disaster recovery stated that disaster recovery must be achieved in an economical manner and that business operations must be restored promptly at the time of a disaster. With the aim of satisfying these requirements, we studied the creation of a new OSS disaster-recovery scheme [2][3].

We studied, in particular, the wide-area, distributed arrangement of a redundantly configured operations system among geographically separated sites (**Figure 8**). The following two requirements arise for achieving such a disaster-recovery scheme.

(1) Location transparency: Applications making up the operations system must be able to run and interact regardless of where those applications might be physically located.

(2) Availability: Business continuity must be achieved at the time of a large-scale disaster through automatic switching to applications at sites not affected by the disaster.



*This is provided only in Japanese at present.

Figure 7 Screen shot of monitoring screen for network control during congestion

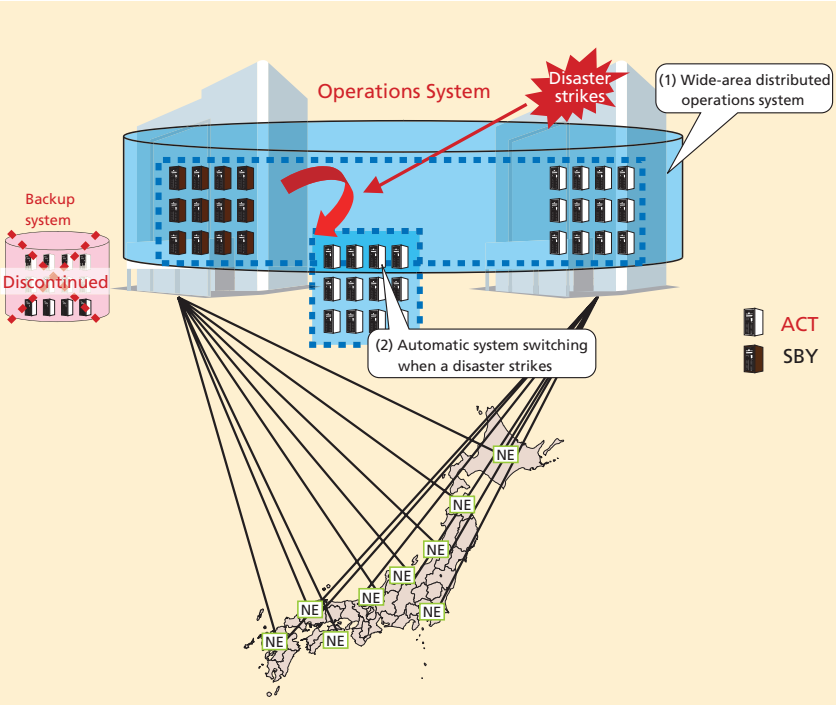


Figure 8 Study of a new disaster-recovery scheme

2) Solution by D3A

We have met the above technical requirements by applying the features of NTT DOCOMO's proprietary Distributed Data Driven Architecture (D3A)[*7] [4].

A system for achieving location transparency in this way is shown in **Figure 9**.

The OSS performs a task by combining multiple OSS applications. To do so, it needs information on which applications to combine and on what order to execute them in. It also needs to know the servers on which those applications are located.

A D3A scenario[*8] defines the applications to combine and the order of their execution in eXtensible Markup Language (XML)[*9].

At the same time, the physical addresses of the servers storing the applications in question are resolved by the Domain Name System (DNS)[*10]. On receiving a D3A scenario, the D3A PlatForm (PF)[*11] section stored on each server requests the DNS to return the IP address of the server storing the next application to be processed.

This system enables applications to run regardless of where they are physically located, that is, regardless of what sites they have been installed at. Location transparency provided by D3A in this way makes it possible to deploy a StandBY (SBY)[*12] redundant configuration for each application at geographically separated sites.

A system for achieving availability is shown in **Figure 10**.

Here, the PF stored on a redundant SBY server periodically accesses the corresponding ACT[*13] to perform a health check[*14]. Then, if a problem is

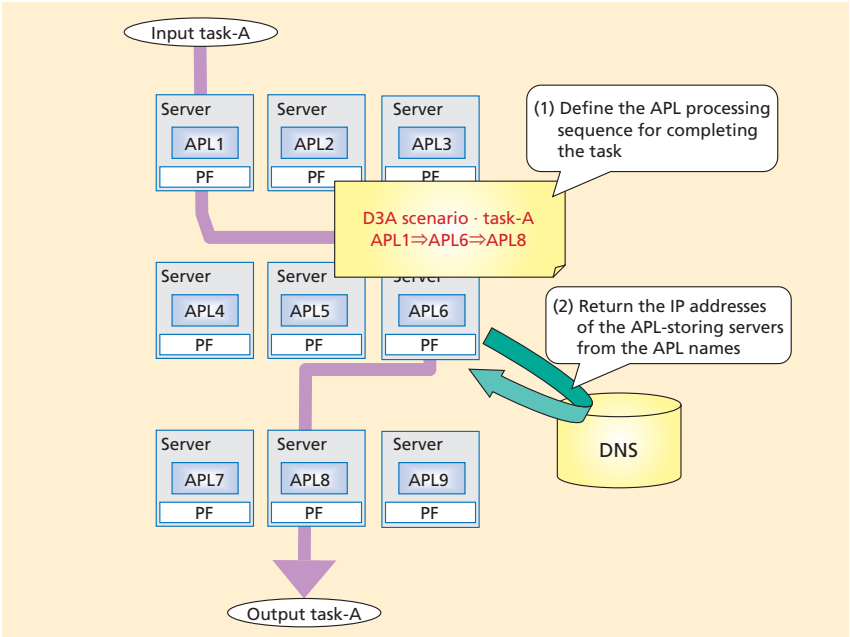


Figure 9 System for achieving location transparency

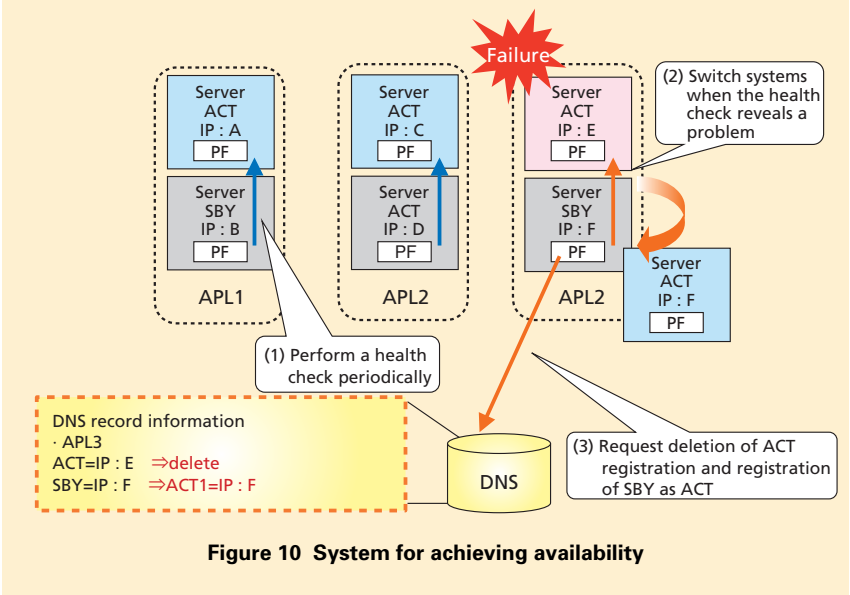


Figure 10 System for achieving availability

*7 **D3A**: An architecture that groups multiple IA servers to enhance processing performance. IA servers use microprocessors from Intel Corp.
*8 **D3A scenario**: A message sent or received by an application running on D3A.
*9 **XML**: A markup language for describing the meaning and structure of text and data.
*10 **DNS**: A system for converting the names of applications running on servers into IP addresses.
*11 **PF**: Refers to the D3A platform section that receives IP addresses from DNS and executes appropriate operations on receiving D3A scenarios.
*12 **SBY**: The parts of a redundant hardware configuration that are currently on standby.
*13 **ACT**: The parts of a redundant hardware configuration that are actually being used.

detected, the PF puts its own server into ACT mode and requests the DBS to delete the previous ACT information and register the new ACT information. The application that had been running on the former ACT server is consequently released from the system and the application running on the new ACT server is incorporated into the system thereby guaranteeing business continuity. In other words, system availability can still be achieved even when servers fail as a result of a disaster at an OSS installation site by having the PF sections of SBY servers automatically switch to applications running at a site unaffected by the disaster.

3) Proposal of a Revised Disaster-recovery Scheme

Putting the "location transparency" and "availability" features of D3A to good use, we proposed a new disaster-recovery scheme as shown in **Figure 11**.

In this scheme, location transparency enables a system configuration that distributes a redundant configuration of the operations system over a wide area at geographically separated sites. This arrangement makes a dedicated backup system unnecessary thereby reducing facility costs.

The availability feature of D3A, meanwhile, makes it possible to automatically switch operations from an OSS installation site affected by a disaster to an OSS application group stored on servers at an unaffected site.

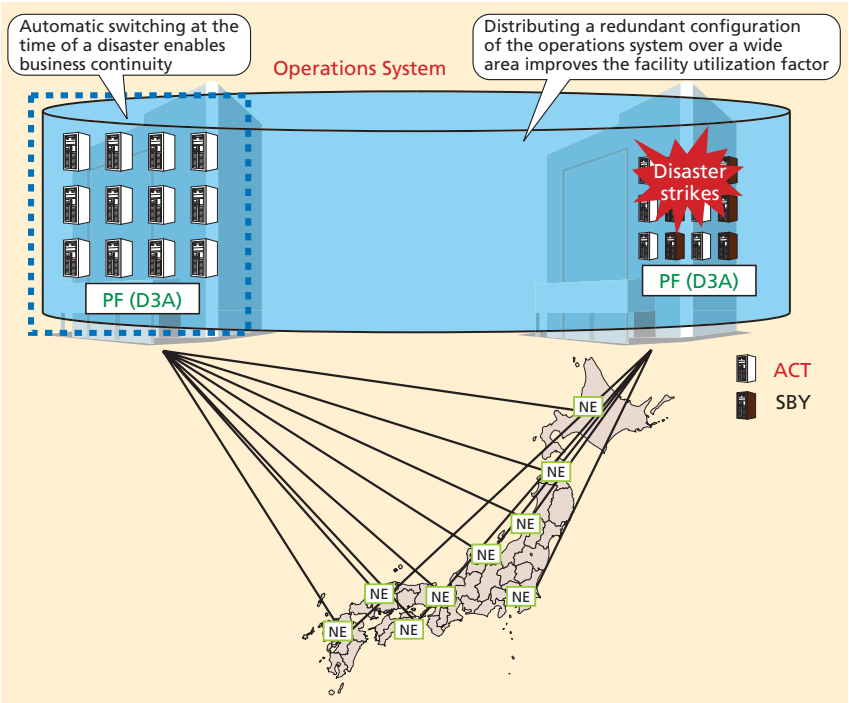


Figure 11 New disaster-recovery scheme

Compared to the former backup system in which some system switching had to be performed manually, this new scheme can shorten the time required for restoring business operations.

- Problems associated with revised disaster-recovery scheme

  The following two problems associated with the revised disaster-recovery scheme had to be solved:

  (a) Since the revised scheme distributes a redundant configuration of the operations system over a wide area at geographically separated sites, transmission-path delays can affect inter-application communications and degrade processing performance.

  (b) Many applications will switch systems when an OSS installation site falls victim to a disaster, but since some applications will have a dependency relationship with others, the order of application system switching must be controlled.

- Countermeasures to problems

  (a) Countermeasure to transmission-path delays

  The noticeable deterioration in performance caused by transmission-path delays in OSS inter-application communications results from the use of synchronous communications. This type of communications guarantees order in processing:

*14 **Health check**: A method for detecting abnormalities by having equipment periodically check the operational state of neighboring equipment.

any one process in a sequence of processes will wait for the previous process to complete before executing. As a result, transmission-path delays will only add to the response-wait times from the message-receive side and degrade performance (**Figure 12**).

As a countermeasure to such transmission-path delays, we added processing that handles inter-application communications as a pseudo-asynchronous sequence making use of D3A PFs. An overview of this countermeasure is provided in **Figure 13** and summarized below.

(1) On receiving a send request from an application, the PF on that server immediately places that request in a queue and returns a response. It then prompts the application for a send request corresponding to the next stage of processing.

(2) Once the created queue has been in existence for a certain amount of time or reached a certain number of scenarios, the PF transmits all scenarios stored in the queue in batch taking their order into account.

(3) The receive-side PF analyzes the scenarios in order and hands over processing to the application in that order thereby guaranteeing ordered scenario processing and avoiding the effects of transmission-path delays.

(b) Controlling startup order in system switching

When applications undergo system switching, communications will occur between some functions that hold important information for executing applications, such as the function that manages the state of system operations and the function that manages the configuration of NEs targeted for monitoring. At the time of system switching caused by a disaster at an OSS installation site, the switching of the servers that hold this impor-

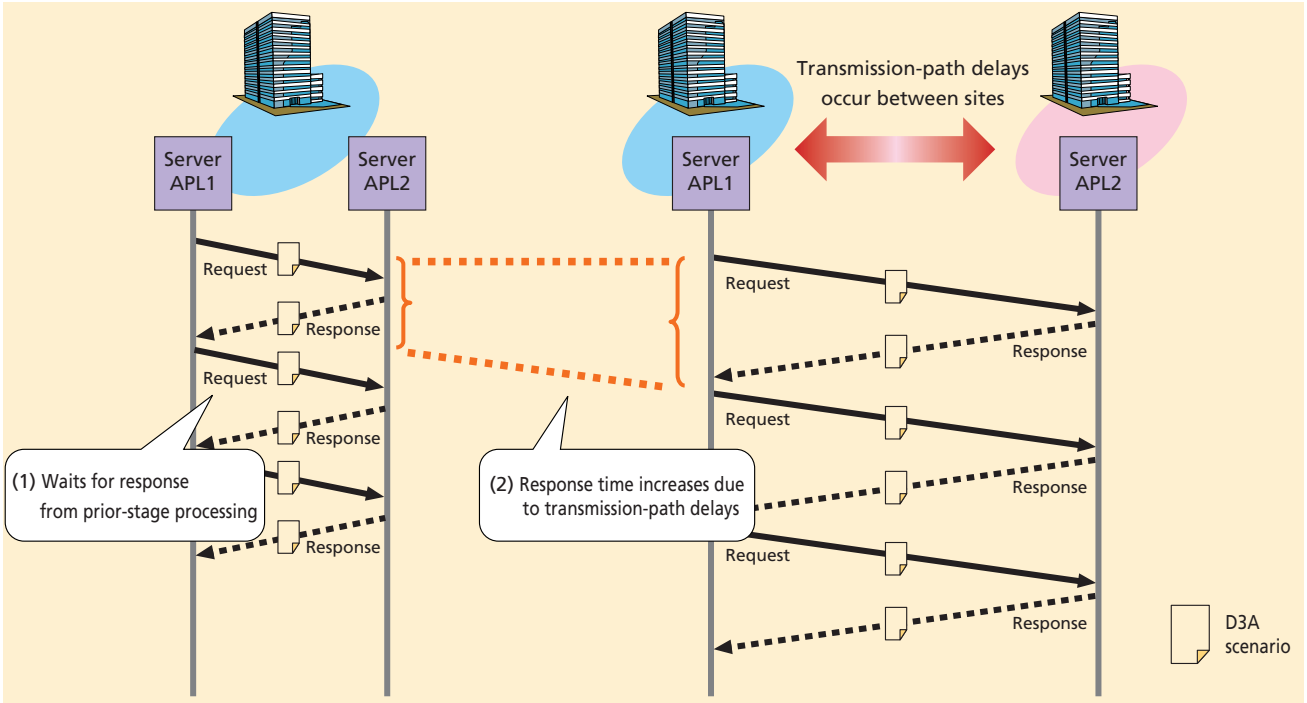


Figure 12 Effects of transmission-path delays in synchronous communications

tant information must absolutely be completed before starting up the switched applications.

The D3A PF of each server is used to control the order of application system switching (**Figure 14**). Specifically, the heath check function provided by the PF of each server is used to check for abnormalities in each application as usual but the timing for carrying out that check is fine-tuned to control

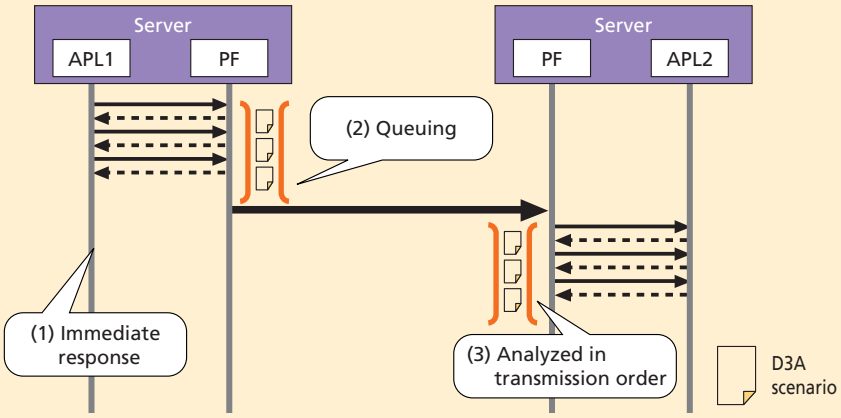


Figure 13 Countermeasure to transmission-path delays by a pseudo-asynchronous sequence

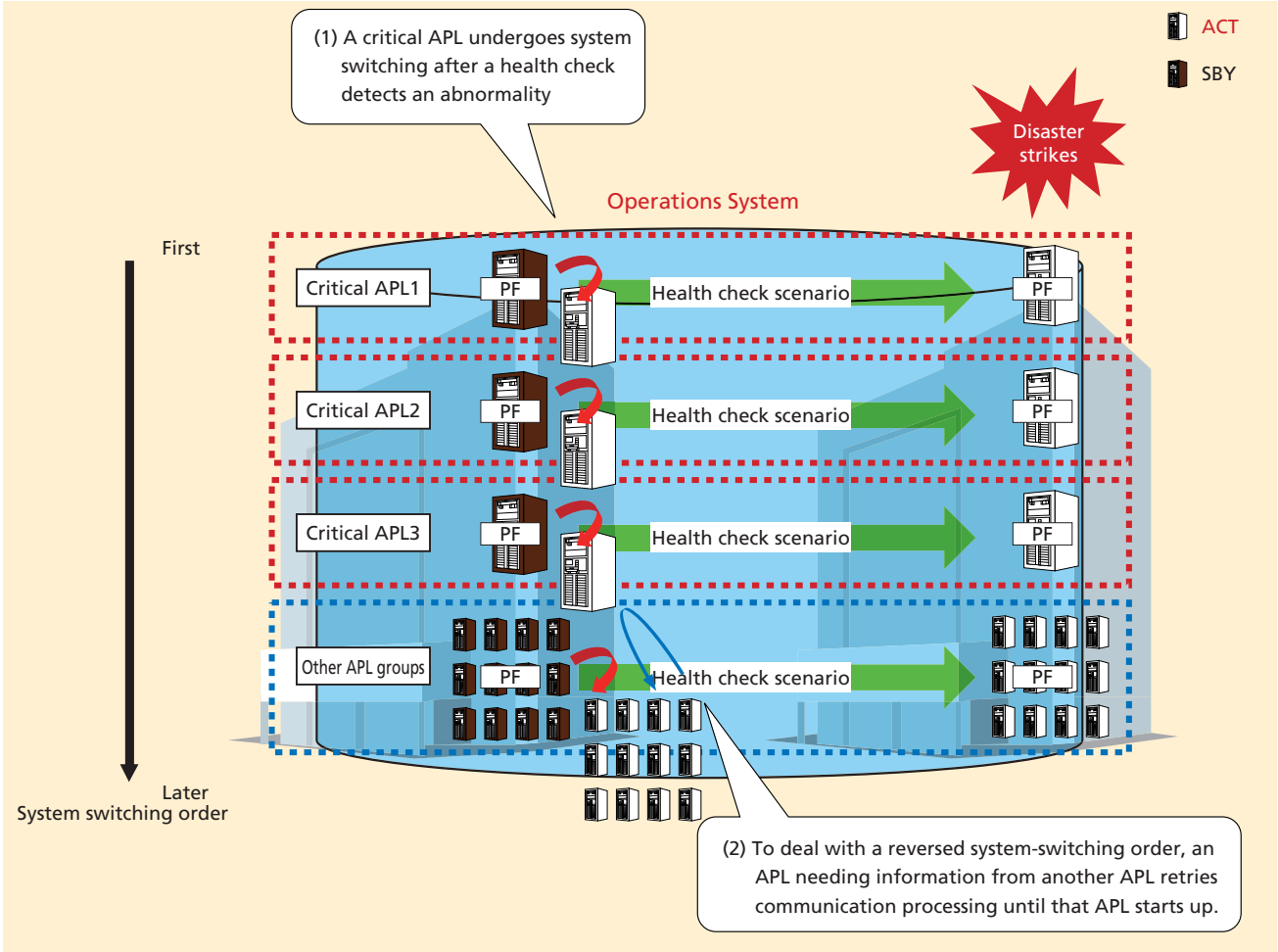


Figure 14 Measure for controlling startup order

the order of system switching. Furthermore, in the event that system-switching order somehow turned out to be reversed, an application that needed information for startup purposes from another application that had not yet started up is made to repeat (retry) the communication process for obtaining that information until that application starts up. The above approach makes it possible to complete application startup correctly. It guarantees accurate system switching of many applications if an OSS installation site should be damaged by a disaster and provides for the continuity of system operations.

The above countermeasures have made it possible for us to implement our new disaster-recovery scheme and thereby achieve economical disaster recovery and a system robust to disaster.

4. Conclusion

Implementing the countermeasures introduced in this article in NTT DOCOMO's OSS have made it possible to determine network damage conditions even at the time of a large-scale disaster. They have also helped to make the OSS itself stronger in the face of congestion and physical damage so as to achieve a system robust to disasters and capable of business continuity.

These countermeasures have therefore helped NTT DOCOMO fulfill its mission as a telecommunication carrier to provide stable network quality at all times.

The revised disaster-recovery scheme described in this article is now being applied to OSSs targeting NEs in the core system, access system and link system. Good results in achieving stable operation are being reported. Looking forward, we plan to study the introduction of this revised scheme to OSSs for other categories of NEs.

REFERENCES

[1] H. Takada, H. Tamura, M. Furutani and K. Takahashi: "Proposal of a system that monitors network elements in the event of a large scale disaster," IEICE Technical Report, Vol. 111, No. 488, pp. 7-12, Mar. 2012 (in Japanese).

[2] Y. Takeuchi, H. Tamura, K. Takahashi, K. Yoshimura and Y. Suzuki: "Challenges to "Location Free" of Distributed Data Driven Architecture (D3A): A study on macroscale disaster recovery," IEICE Technical Report, Vol. 110, No. 24, pp. 11-16, May 2010 (in Japanese).

[3] Y. Takeuchi, K. Kagawa, H. Tamura, M. Furutani and K. Takahashi: "The Practical Application of Geographically Distributed OSS," IEICE Technical Report, Vol. 112, No. 120, pp. 25-30, Jul. 2012 (in Japanese).

[4] K. Akiyama et al.: "Technology for Achieving an Economical Operations System—Distributed Data Driven Architecture—," NTT DOCOMO Technical Journal, Vol. 13, No. 2, pp. 36-46, Jul. 2005 (in Japanese).