Restoring CSCF by Leveraging Feature of Retransmission Mechanism in Session Initiation Protocol

Takeshi Usui and Nozomu Nishinaga
New Generation Network Laboratory
National Institute of Information and Communications Technology
Koganei, Tokyo
e-mail: ta-usui@nict.go.jp and nisinaga@nict.go.jp

Yoshinori Kitatsuji and Hidetoshi Yokota
Mobile Network Laboratory
KDDI R&D Laboratories, Inc.
Fujimino, Saitama
e-mail: kitaji@kddilabs.jp and yokota@nict.go.jp

Abstract— **The IP multimedia subsystem (IMS) is a key technology for providing various services over IP-based networks. IMS enables network service providers to collect information relating to the communications of customers, such as an accounting through call/session control function (CSCF) which is used to establish a session using the session initiation protocol (SIP). Therefore, the availability of CSCF has become important. We propose a system to recover the session states maintained by the CSCF in the alternate CSCFs. This is achieved in a low cost solution by leveraging the features of retransmission mechanism in SIP. Our proposed system selectively saves the session state in order to reduce the saved data and recovers the saved session state in the alternate CSCFs rapidly when the faults occur in CSCFs. We show that our proposed system can achieve 60% reduction of the backup servers and that the overhead of our proposed system is not large.**

Keywords-component; Restoration system; IMS; SIP; Network Operation

I. Introduction

Many network service providers (NSPs) are now promoting convergence towards providing various services over IP-based networks. IP Multimedia Subsystem (IMS) [1] is supposed as the service management and control function over IP-based networks. In the IMS, a call/session control function (CSCF) [2], that is a session initiation protocol (SIP) [3] proxy server, is used to establish a session, that accompanies the information to start the application between user equipments (UEs), through the SIP. When the UE starts an application managed by IMS, such as IP telephony, a series of SIP messages are exchanged through multiple types of CSCFs that is referred to as the SIP signaling call flow.

The availability of the CSCFs is essential for session management, i.e., service control, administration, and accounting. If the CSCFs are not available, the UEs cannot start the application. In the worst case, the services managed by IMS may go down because the application initiation and the communications of the UE are broken off. Therefore, the availability of the CSCFs is of great importance.

For continuing services, the service status need to be recovered to an available server and the service execution is succeeded with this server when the original server becomes unavailable. E.g., in the case of the CSCFs, it is required to recover the session states which the CSCF keeps for each UE. A session state is the information which is used for establishing the session (i.e., transaction and dialog, details are described in Section III-B). For restoring the CSCF, the NSP may adopt a fault tolerant (FT) server [4] which consists of a pair of active and backup hardware and copies all the data from the active server to the backup server. In this case, the NSPs must prepare the same number of backup servers as active servers. As the other methods, Peer-to-Peer (P2P) SIP [5] can be used to improve the redundancy of SIP proxy servers. However, P2P SIP cannot recover a session state when the SIP signaling call flow is not completed and a session is not yet established (hereafter, this incomplete session is termed “pre-session”). Here, we propose a system that recovers even the pre-session as well as the established sessions from the aggregated backup servers.

For completely restore the CSCF that halted because of faults (e.g., hardware fault) (hereafter, termed “fault CSCF”), the session states must be saved in the backup server which is positioned as the backup hardware in the case of FT and the alternate CSCF needs to take over the session state of the fault CSCF as described in a literature [6]. However, this process has two issues. The first is that the NSP encounters to maintain a number of the backup servers because of the high volume of the data to be saved. Now, NSPs aim to reduce the number of the servers for cost saving by aggregating multiple servers to fewer servers.The second is that the NSP needs to restore the CSCF within a limited period. This is because the NSP never want to lose the information related to the accounting.

To solve the first issue, we reduce the number of the backup server by selectively saving the session states. The SIP application retransmits SIP messages if they are lost. Our proposed system does not save the session states that can be restored by the retransmitted SIP messages, but save the session states based on the relationships between the retransmitted SIP messages and the session state that is necessary for the CSCFs to handle the retransmitted SIP messages. Our proposed system enables few backup servers to save the session states from a large number of CSCFs because the volume of the saved data is sufficiently reduced.

To solve the second issue, we propose that the specific session state be recovered rapidly in turn based on the type of each session state. We defined priorities from the session states which are required from the service. At first our proposed system recovers the pre-session in order to

 ISBN: 978-1-61208-174-8

continue the SIP signaling call flow and restricts the influence of the CSCF faults at the minimum level.

The rest of the paper is organized as follows. Section II and III describe the requirement for CSCF restoration system and the design of our proposed system. Section IV evaluates the overhead of our proposed system and how many backup servers are reduced. Section V concludes the paper.

II. REQUIREMENTS FOR CSCF RESTORATION SYSTEM

A. Overview of IMS Architecture

One of the features in the IMS is that three different SIP proxy servers are used. The core functional components of the IMS are these three different kinds of CSCFs: Proxy-CSCF (P-CSCF), Interrogating-CSCF (I-CSCF), and Serving-CSCF (S-CSCF). These components are responsible for the management of SIP signaling call flow, such as routing a SIP message, and authentication and registration for UEs.

The P-CSCF is the first functional point of contact between UE and IMS components. It maintains IPSec security associations (SA) with UE and applies integrity and confidentiality protection for the SIP message. The I-CSCF is responsible for assigning a suitable S-CSCF to the UE and routes incoming requests to S-CSCF. The S-CSCF handles the registration process and downloads authentication data from the Home Subscriber Server (HSS), which is a database containing the information base of the subscribed users, such as the users identifiers, location information, and so on.

In this paper, we focus on the P-CSCF and the S-CSCF for the restoration system because these servers maintain the sessions in a stateful mannr. Because the I-CSCF deals with the SIP messages in a stateless manner, the UE and CSCFs does not need to execute registration process from the beginning when the alternate I-CSCF is prepared. Therefore, recovering the session states of the I-CSCF is not necessary.

B. Functions of CSCF Restoration System

In this paper, we assume the CSCFs and UE use UDP as the transport protocol for exchanging SIP messages because many NSPs do so. Therefore, the CSCFs and UE use the retransmission mechanism in SIP. (When adopting TCP, our proposed system cannot be used because the CSCFs do not use the retransmission mechanism)

The following three functions are considered necessary for the restoration system.

(1) Monitoring function to detect CSCFs failure quickly
(2) Saving function to save the minimized data of the session states from the CSCFs before their faults occurred
(3) Restoring function to rapidly recover the session state in the alternate CSCF without requiring any action of UE or without losing session management at the UE due to the faults.

After function (1) detects faults of CSCF, we can execute the procedures to recover the session states. In this paper, we assume that the existing tools and method (e.g., [7] [8]) are used to detect the faults of CSCF. We thus focus instead on the design and implementations of functions (2) and (3) in a low cost solution.

In function (2), a key is reducing the saved data of the session state. To implement the low-cost restoration system, it is desirable that the volume of the saved data is low. Saving all the change of session state as the CSCF receives any types of SIP messages leads to a large amount of data transferred to and stored in the storage. The reduction of the saved data allows the number of backup servers maintained in the NSP to be reduced. We distinguish SIP message (INVITE, UPDATE, ACK, and so on) based on the priorities in recover the session state just after the CSCF is halt, and let the session state be saved when the important SIP message is received by the CSCF. Thus, the CSCF saves the limited session states instead of saving every change of the session state. The definition of the importance of SIP message and the other detail are described in Section IV.

Function (3) needs to be executed as rapidly as possible in order to not affect the recovered sessions maintained in the alternate CSCF server after the failure of the original CSCF server. No particular functions in the implementation of the UE should be additionally required for function (3).

Function (3) is also required to recover the pre-session. If the pre-session is not recovered, the UE fail to start the application and needs to execute the application initiation procedures from the beginning. In the SIP, there is no procedure to recover the broken procedure of the application initiation. In order that the UE start to use the application, that procedure is need to be defined.

C. Related Work

The FT server [4] requires the pairs of active and backup hardware resources because it saves data from active server to backup server whenever the CPU executes any commands. Therefore the FT server requires the specialized hardware is required. The FT server has function (3), but does not satisfy function (2).

The redundancy techniques for the web servers [9] [10] are applicable for CSCFs. However, these technologies do not achieve minimizing the saved data of the session states because these technologies do not consider the feature of the SIP. These technologies do not satisfy function (2).

The system for the replication of the SIP proxy server as described in [6] has been proposed. In this system, a pair of an active and backup server is prepared. The replication is executed every time SIP messages are exchanged between the SIP proxy and the UE. This system satisfies the function (3). But, this system does not consider the reduction of the data to be saved in the backup servers. Therefore, this system does not satisfy the function (2).

P2P SIP is also used to improve the redundancy of the servers, but it cannot recover the pre-session because there is no mechanism that shares the session states that the SIP proxy servers keep. For P2P SIP, it is necessary for the UE to re-register with the SIP proxy server so that a different SIP proxy server manages the session state of the UE. P2P SIP does not satisfy the function (3) because the data for restoring the CSCF is not saved anywhere.

 ISBN: 978-1-61208-174-8

III. Proposed Restoration System

This section presents the design of our proposed system which leverages the features of the retransmission mechanism in SIP.

A. *Overview of our proposed system*

An overview of our proposed system is shown in Figure 1. The restoration system consists of multiple active and backup servers and does not include the monitoring function. The backup servers save the session states from multiple CSCFs through the out-band line (bidirectional arrowed lines in Figure 1). When the monitoring system detects a fault, the restoration system is commanded to set up an alternate CSCF. The alternate CSCF is selected from one of standby servers and takes over the IP address and configuration of the fault CSCF. The alternate CSCS is prepared before the migration of the transactions and dialogs.

Our proposed system takes the following steps if the CSCF halts with the fault.

1. The monitoring system or the operator order the restoration system to restore the CSCF
2. The alternate CSCF is set up and the saved transactions and dialogs are recovered from the backup server in order to take over the session states.

B. *Analsys of SIP Signaling Call Flow*

The registration procedure from UE-A is shown in Figure 2. The procedure is normally executed when UE is turned on. In this call flow, authentication and key agreement (AKA) [12] that is necessary for the UE to get access the IMS-based services is also executed, and the IPSec is established between the UE and the P-CSCF. If any SIP messages are lost, the first Register message is retransmitted.

We assumed the application initiation procedures of IMS as shown in Figure 3. Each SIP message in the Figure is numbered. The INVITE message is used to initiate a session between UE-A and UE-B. Here, UE-A and UE-B register with the different P-CSCFs and S-CSCFs. We next summarize the retransmission behavior of the SIP messages by separating the SIP signaling call flow into seven phases. The SIP message numbers are in parentheses.

In Figure 3, there is a Request-Response relationship among the SIP messages, such as Bye and 200OK. In the case that a SIP message is lost, the SIP signaling call flow always restarts from the SIP messages that corresponds to the request message. The CSCF forwards an incoming request message to the UE or the other CSCF and keeps waiting for the response message until either the response message is received for the request message or a set timer expires. This relationship among the SIP messages is called a “transaction” in SIP. A “dialog” means the state that the CSCFs create about the communications between the UEs through the SIP signaling call flow. The dialog is kept in the CSCFs while the UEs communicate, and erased when the communication is terminated.

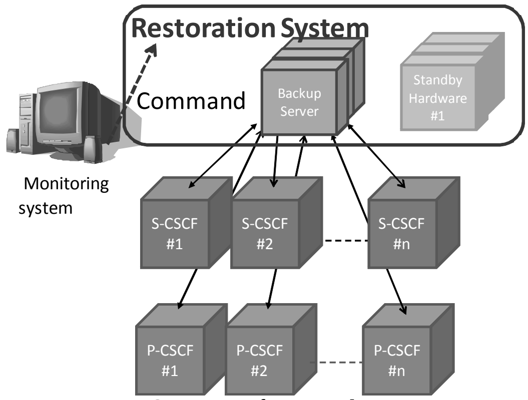

Figure 1 Overview of proposed restoration system

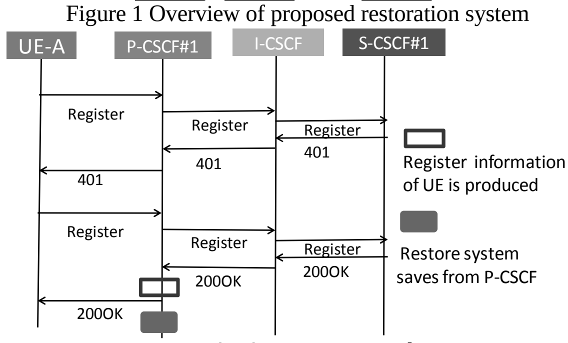

Figure 2 AKA Register procedure

In this paper, we define that the session state consists of the transaction and dialog. A transaction and dialog are updated whenever SIP messages are exchanged through the SIP signaling call flow. The specific transaction and dialog during the SIP signaling call flow are required in order to recover the pre-session in the alternate CSCF. When the CSCF does not have the transaction and dialog related to that SIP message, the CSCF cannot handle that SIP message. Our proposed system save the session states that are required to recover the pre-session.

1) INVITE/100Trying/183Session Progress (1–17)

This phase is initiated with the sending of an INVITE message and is terminated when UE-B receives 100Trying or 200OK message. This phase actually consists of multiple parts: part of INVITE and 100Trying message and part of INVITE and 183Session Progress. Once the P-CSCF#1 receives the INVITE message from UE-A, it replies with 100Trying message to UE-A. The same exchange of the SIP messages applies to each hop until the INVITE message reaches UE-B. If the INVITE or 100Trying message is lost, the UE will retransmit the request after the expiration of a timeout value called T1. If the INVITE message sent by the P-CSCF to the S-CSCF or the 100Trying message sent by the S-CSCF to the P-CSCF is lost, the P-CSCF will retransmit the INVITE message after T1 seconds. The same applies to all hops until UE-B. Note that the INVITE message is only the case in which the reply message are sent

 ISBN: 978-1-61208-174-8

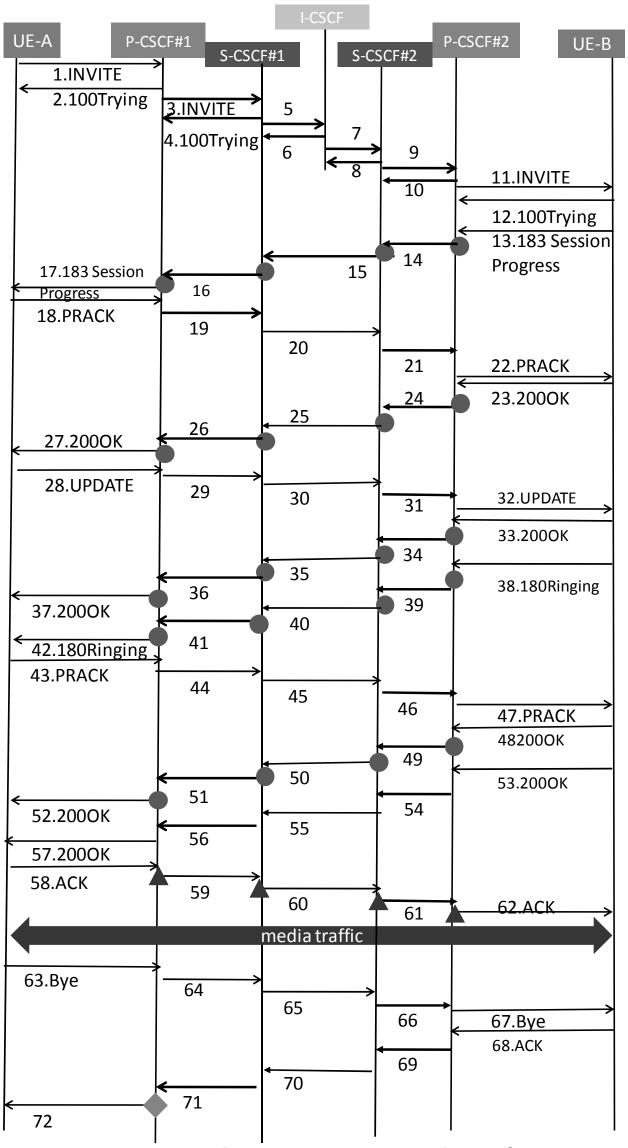

Figure 3 Application initiation procedures of IMS

hop-by-hop. In all the other phases 2) through 7), the UE is responsible for generating appropriate acknowledgements in an end-to-end manner. In the rest of this phase, UE-A and UE-B exchange the information of the audio and video codes to be used as well as the quality of service (QoS) criteria. The 183Session Progress message is sent reliably. If this message is lost, the INVITE message is retransmitted.

2) PRACK/200OK (18–27)

UE-B acknowledges receipt of the 183Session Progress message. The PRACK message will be retransmitted by UE-A until a 200OK message is received. Note that the transaction related to the PRACK message starts its retransmission timer just after it receives the 183Session Progress message. Furthermore, the PRACK message will not be retransmitted each time the 183Session Progress message is retransmitted but only when a retransmission timer is triggered (i.e., T1, 2T1, etc.).

3) UPDATE/200OK (28–37)

UE-A and UE-B complete the exchange of the code and QoS information. If the UPDATE message or the 200OK message is lost, the caller will retransmit the UPDATE message until the 200OK message is received.

4) 180Ringing/PRACK (38–47)

UE-A informs the caller that the user is being alerted about a call. The 180Ringing message is sent reliably. That is, UE-A retransmit it until the PRACK message is received.

5) PRACK/200OK (43–52)

UE-B acknowledges receipt of the 180Ringing message. The PRACK message will be retransmitted by UE-A until a 200OK message is received.

6) 200OK/ACK (53–62)

The callee informs the caller that the call was accepted. If the response or an ACK message is lost, the callee will retransmit the 200OK message until the ACK message is received.

7) BYE/200OK (63–72)

One UE terminates the call. If the BYE or acknowledging 200OK message is lost, the sender of the BYE message will retransmit that message until a 200OK message is received.

C. Saving Session States from CSCFs

Saving all the transactions and dialogs in the SIP signaling call flow generates a number of request messages and its acknowledgement messages for saving the session states and makes the load of the backup servers high. Therefore, a certain number of backup servers are necessary for the restoration system. When the alternate CSCF is set up and the UE sent one SIP message in the SIP signaling call flow, the CSCF cannot handle the SIP message if the CSCF does not have the transaction and dialog related to the SIP messages that UE sent.

The backup servers do not need to save the session states whenever the CSCF sends the SIP messages. Saving the specific session states that is necessary for the CSCF to handle the retransmitted SIP messages is enough. To reduce the load of the backup servers, we propose that the restoration system not save the transactions and dialogs that are re-generated after the retransmission by the UEs. We select the specific session states by leveraging the features of SIP retransmission mechanism.

Our proposed system save the session states which the CSCFs need in order to handle the retransmitted SIP messages. To recover the session states completely, the SIP signaling call flow is made to wait for the completion of saving from the CSCFs to the backup server. Only after the session states are saved, the CSCF can send the next SIP message in the SIP signaling call flow. If the saving process is not completed and the next SIP messages have been sent before a fault occurs, the UE cannot complete the SIP signaling call flow because the CSCF cannot handle the SIP messages that the UE sent.

The request messages for saving the session states are sent by the CSCF to the backup server and the acknowledgement message for informing the completion of saving is returned by the backup sever to the CSCF. If much time is spent on

 ISBN: 978-1-61208-174-8

treating the request messages from the CSCFs, the SIP signaling call flow is delayed. To prevent this situation, it is necessary to increase the number of backup server or to reduce the number of the total durations for saving the session states. In Section IV, we evaluate about the saving process in the backup server.

For the registration procedure (Figure 2), the restoration system saves the register information (e.g., UE IP address and URI) and the key information for the IPSec. The rectangle in Figure 2 indicates the points at which the restoration system saves the register and key information.

For the application initiation procedure (Figure 3), the transactions and dialogs which the CSCFs need in order to handle the retransmitted SIP messages are summarized in Table 1. We assign phase numbers for the sequenced SIP messages. The circles in Figure 3 indicate the points at which the restoration system need to save the transactions and dialogs before the CSCFs send the next SIP messages to the UE or the other CSCF.

If the 183Session Progress message does not arrive, or the other SIP messages are lost, the UE retransmits the INVITE messages. For the INVITE, 100Trying, and 183Session Progress messages (phase 1), the restoration system does not need to save any session states. After the 183Session Progress message is sent, the transaction and dialog need to be saved for when the PRACK message or the 200OK message is retransmitted (phase 2). After the 200OK message is sent from UE-B, the transaction and dialog need to be saved for when the UPDATE message or the 200OK message is retransmitted (phase 3). To handle the retransmitted 180Ringing message (phase 4), the CSCFs need to save the transaction and dialog after sending the 200OK message. After the CSCFs send the 180Ringing message, the restoration system needs to save the transaction and dialog for when the PRACK or 200OK message is retransmitted (phase 5).

After the ACK messages are received by the CSCFs (phase 6), only the dialogs are saved from the CSCFs because. The triangles on the SIP signaling call flow in Figure 3 indicate the points at which the restoration system saves only the dialog because the transaction is erased in the CSCF. The session is then established, and the keep-alive messages are exchanged in some periods until one UE sends the Bye message. After a few minutes, the restoration system erases the transaction because the session is established and no retransmission is generated from now. As for the dialog of the Bye and 200OK messages (phase 7), the restoration system erases the dialogs of the UE. The diamond on the SIP signaling call flow in Figure 3 indicates the point at which the restoration system erases the dialogs of P-CSCF#1, S-CSCF#1, S-CSCF#2 and P-CSCF#2.

D. Recovering Session States

We propose that the restoration system puts priorities on the session states in order to recover the session states which will be used by the CSCF immediately because the SIP signaling call flow is processed (not completed) and the UE retransmits the SIP messages. For this prioritization, we classify three types of session states. The session state at which the SIP signaling call flow is not completed between the UEs is “pre-session”, the session state in which the final response 200OK message and ACK message are sent and received is “active session”, and the session state in which the UE only registers with the CSCFs is “register session”.

Table 1 Relationship between retransmitted SIP messages and transactions and dialogs necessary for CSCF to handle the retransmitted SIP messages

Phase No.	SIP message	Transactions and dialogs are saved before SIP proxy server send SIP message
1	1-17	None
2	18-27	14,15,16,17
3	28-37	24,25,26,27
4	38-42	34,35,36,37
5	43-52	39,40,41,42
6	53–62	49,50,51,52

In our proposed system, first the pre-sessions must be recovered because the application initiation by the UE will halt if the CSCFs do not return the SIP messages. Second, the active sessions are migrated and restored. Even if the restoration of the active-sessions is delayed, there is not severely influence on the communication between the UEs. But it is important to manage the termination of the communication for the accounting. Finally, the register sessions are recovered

The number of pre-sessions in the CSCFs is small compared to the register and active sessions, although the CSCFs accommodate a large amount of UE. Therefore, we consider that our proposed system enables the transactions and dialogs to be migrated into the alternate CSCF and the alternate CSCF to conduct the session management on them at an earlier time in the entire session restoration.

Because the active sessions are migrated second, a probability exists that the Bye messages are sent by the UE before migration of the active sessions is completed. In this case, the CSCFs normally return error code messages and the some applications of the UE stop retransmitting the Bye messages and terminated, which depend on the implementation. As a result, the NSPs then cannot obtain the accounting information from the UE. To solve this problem, we append a function, that is, the CSCF discards the Bye message and does not return any error messages before the completion of restoring the session state. When the UE retransmits the Bye messages, the session is terminated as normal by the CSCF, which takes over the session state.

IV. EVALUATION

A. Model of Experiment

Our proposed system aims to reduce the number of backup servers. The number of CSCFs treated by the single backup server increases as the number of backup servers decreases. In the case that the process for saving session state congests in the backup server, the slow responses from the backup server to the CSCF affects the completion time of the SIP signaling call flow. We define the queuing delay as the duration taken by the backup servers to finish the process for

 ISBN: 978-1-61208-174-8

Table 2 Parameters employed in experiments

Propagation Delay	Value (millisecond)
Between UE and P-CSCF	10
Between CSCFs (P-CSCF#1 and S-CSCF#1,S-CSCF#1 and S-CSCF#2, S-CSCF#2 and P-CSCF#1)	1
Between backup server and CSCF	10

saving the session states. If a large queuing delay for treating the request messages that arrive continuously from the CSCFs occurred in the backup server, the duration for establishing the session between the UEs is delayed because the CSCF can send the next SIP message in the SIP signaling call flow only after the session states are saved. The approximate propagation delay is easy to estimate beforehand between the backup servers and the CSCFs because NSP can constructs dedicated network for the transaction between CSCFs and backup servers and the sufficient network capacity can be prepared. We evaluated how long the queuing delay in the backup servers will affect on the duration for establishing the session between the UEs with an event-based simulator.

The CDMA 2000 defines about 5 seconds as the suggested time of paging process [13]. This indicates that the additional duration for establishing the sessions between the UEs is sometimes required when the location of the UE is searched for. We adopt 5 seconds as the criterion in order to evaluate the delayed time of the application initiation procedures by our proposed system. We believe that there is no problem if the overhead of saving session state reached to the same as the duration of the paging process and that total duration for establishing the session is not beyond 20 seconds even if the paging process spends about 5 seconds.

We compared our proposed system to the case where the backup servers save the session states every time the P-CSCFs and S-CSCFs send SIP messages (hereafter, termed “Allcopy”), that is same with the approach reported in the [6]. In this evaluation, the number of backup servers (= *b*) was changed and Figure 2 is used as the SIP signaling call flow where 2 set of P-CSCFs and S-CSCFs exist.

We also assumed that 100 million UEs are accommodated by multiple CSCFs. The call arrival rate follows the Poisson process for that number of UE. Additionally, to represent an on-peak period, the offered calls were 20 times larger than the 1-year average call arrival rate (1.7 calls per user a day) [13]. The duration of call ringing was a uniform random numbers between 1 and 5 seconds. The call duration followed the exponential distribution with 120 seconds as the average [13].

The parameters for the for the propagation delay of links are shown in Table 2. We used 10 and 20 microseconds, respectively as the time *c* for executing saving the session states in the backup server after the backup servers receive the request messages for saving from the CSCFs.

B. Evaluation Result

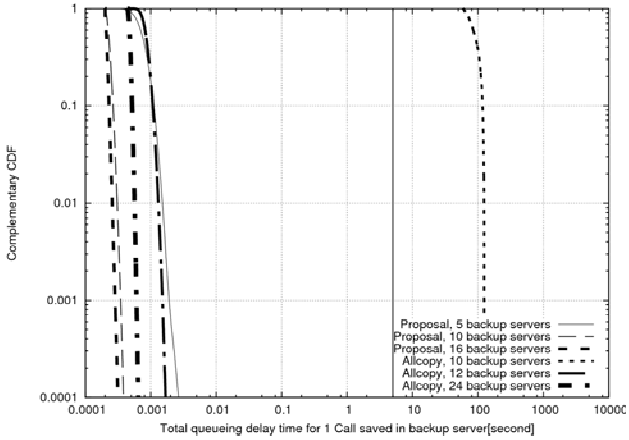

Figure 4 Total queuing delay time for 1 call saved in backup servers when *c*=10 microseconds

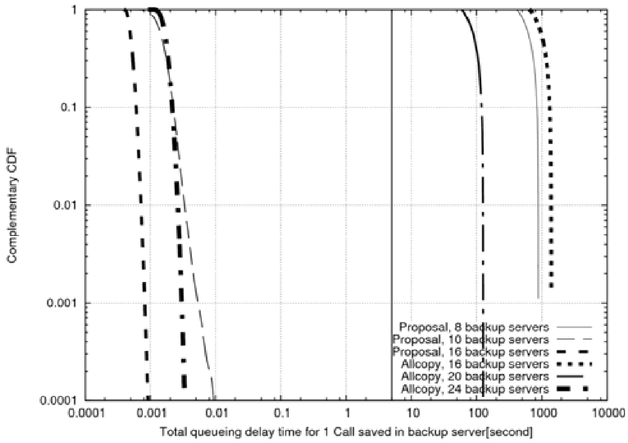

Figure 5 Total queuing delay time for 1 call saved in backup servers when *c*=20 microseconds

As the total number, 72 messages in Figure 3 were processed through the CSCFs, 12 of which are sent by the UEs. In our proposed system, the session states are saved at 20 entry points, at which the CSCFs are made to wait for the completion of saving. The total duration of the propagation delay for sending the session states to the backup server and returning the acknowledgement from the backup server was 400 milliseconds.

In the all copy case, the session states are saved at 48 times from the P-CSCF and S-CSCF (except the cases where the I-CSCF sends the SIP message and where the UE sends the Bye message). In this case, the total time of the propagation delay for sending the session states to the backup server and returning the acknowledgement from the backup server was 960 milliseconds. We consider neither value had a large impact on the UE.

The total queuing delay time for one call (from INVITE message to ACK message) saved in the backup servers for *c* = 10 and 20 microseconds is shown in Figures 4 and 5, respectively. The x-axis represents that the queuing delay in the case *c*= 10 or 20 microseconds but does not include the duration for sending the session states to the backup server

 ISBN: 978-1-61208-174-8

and returning the acknowledgement from the backup server. The y-axis represents the complementary cumulative distribution function (CCDF). Both axes use a log scale. We drew the line which means 5 seconds in Figures 4 and 5.

As Figure 4 shows, our proposed method did not accumulate large queuing delay when *b* = 6, 10 and 16, and *c*= 10 microseconds and gave less impact on the duration for establishing the sessions between the UEs. However, the All copy case accumulated the large queuing delay (more than 5 seconds) when *b* = 10 because the treatment for the request messages in the backup servers was congested. When *b*= 10, the quality of the IMS-based services is degraded because the duration for establishing the session between the UEs is beyond 5 seconds. In this situation, in order not to delay the duration for establishing the sessions, 12 backup servers are necessary for the Allcopy, however 5 backup servers are necessary for our proposed system. Our proposed system achieves about 60% reduction of the backup servers.

As Figure 5 shows, our proposed system did not accumulate large queuing delay when *b* = 10 and 16, and still gave less impact on the duration for establishing the sessions between the UEs. However, our proposed system generated a large queuing delay when *b* = 8. The Allcopy also accumulated a large queuing delay when *b*=16, and 20. This indicates that more backup servers are necessary to shorten the duration for establishing the sessions between the UEs. In this situation, in order not to delay the duration for establishing the sessions, 24 backup servers are necessary for the Allcopy, however 10 backup servers are necessary for our proposed system. Our proposed system also achieves about 60% reduction of the backup servers.

Compared the result in Figure 4 to the one in Figure 5 when *b* =10, 16 and 24, each increased time of queuing delay in the Allcopy is much more than in our proposed system. When the duration for executing saving the session states becomes longer, the effect of our proposed system becomes larger.

V. CONCLUSION

This paper introduced an approach to save and recover the session states of the CSCFs in a low cost solution. In this paper, we aim to reduce the number of the times to be saved by leveraging the feature of the retransmission in SIP. Our proposed system contributes to the recovery of even the pre-session and, at the same time, the reduction of the number of the backup servers. We presented selectively saving and restoring in turn based on the type of the session state. We highlighted the relationship between the retransmitted SIP messages and the transactions and dialogs necessary for CSCF to handle the retransmitted SIP messages. The existing system unnecessarily saves the data from the SIP proxy servers for restoring the servers.

We evaluated the queuing delay which the backup servers spend for saving the session state as the overhead of our proposed system. The overhead of our proposed system was shown with simulation experiments compared to the case where the backup servers save the session states every time the P-CSCFs and S-CSCFs send SIP messages The experiments showed that the overhead of the selectively-saving in our proposed system did not affect much more on the completion of the SIP signaling call flow than the system described in [6]. Our proposed system can achieve about 60% reduction of the backup servers. Our proposed system contributes to cost saving because it succeeds in saving the session states in fewer backup servers.

As the future work, we need to evaluate how long it will take to recover the session states in the alternate CSCF with the implementation of our proposed system.

REFERENCES

[1] ETSI “Telecommunications and Internet converged Services and Protocols for Advanced Networking (TISPAN)”, ETSI E2 282 001 V1.1.1 NGN Functional Architecture Release 1.

[2] ETSI, “Telecommunications and Internet Converged Services and Protocols for Advanced Networking (TISPAN); IP Multimedia Call Control Protocol based on Session Initiation Protocol (SIP) and Session Description Protocol (SDP) Stage 3”, ETSI ES 283 003, V1.1.1, April 2006.

[3] M. Handley, H. Schulzrinne, E. Schooler and J. Rosenberg, "SIP: Session Initiation Protocol", IETF RFC 3261, June 2002

[4] D.Jewett: “Integrity S2: A Fault-Tolerant Unix Platform”, IEEE 21th FTCS International Symposium 1991, pp.512-519

[5] D.Bryan, P.Matthews, E.Shim, D.Willis, and S.Dawkins, “Concepts and Terminology for Peer to Peer SIP”, draft-ietf-p2psip-concepts-03, Oct, 2010

[6] G.Kambourakis, D.Geneiatakis, S.Gritzalis, C.Lambrinoudakis, T.Dagiuklas, and J. Fiedler, “High Availability for SIP: Solutions and Real-Time Measurement Performance Evaluation”, International Journal of Dissaster Recovery and Business Continuity, Vol. 1, No.1, Feburary,2010

[7] A.Kobayashi, H.Ishizuka, N.Tomoed, T.Sone, E.Kosugi, and A.Iwatani, “VoIP network monitoring solution for IP telephony service as public communication infrastructure”, Mitsubishi Electoronics Technical Paper, April, 2006

[8] T.Usui, T.Kubo, Y.Kitaji and H.Yokota, “A Study on Locating Lossy Links of Signaling Messages in SIP-based Services”, IEICE trans-b, vol.E94-B, No.1, January 2011

[9] O. Damani, P. Chung, Y. Huang, C. Kintala, and Y. Wang,"ONE-IP: techniques for hosting a service on a cluster of machines", Computer Networks, vol. 29, pp. 1019-1027, September, 1997

[10] D. Oppenheimer, A. Ganapathi, and D. Patterson,"Why do internet services fail, and what can be done about it?", in 4th USENIX Symposium on Internet Technologies and Systems (USITS '03), (Seattle, WA), March, 2003

[11] 3GPP TS 33.203 (v7.6.0), "3G security; Access security for IP based services", Release 7, June, 2006

[12] V. Ramaswamy and J.Chung, “Performance Analysis of the Quick Idle State Protocol of CDMA 1xEV-Do Rev.B systems”, IEEE Globecom 2010

[13] Telecom Data Book 2010, Telecommuncations Carries Association in Japan, http://www.tca.or.jp/databook/index.html

 ISBN: 978-1-61208-174-8