Distributed Management Architecture for Multimedia Conferencing Using SIP

Yeong-Hun Cho[1], Moon-Sang Jeong[2], Jong-Tae Park[2], Wee-Hyuk Lee[2]
[1] *Department of Information and Communication, Kyungpook National University*
[2] *School of Electrical Engineering and Computer Science, Kyungpook National University*
1370, Sankyuk-Dong, Buk-Gu, Daegu, Korea 702-701
{yhcho, msjeong, jtpark, whlee}@ee.knu.ac,kr

Abstract

As various multimedia communication services are increasingly required by Internet users, several signaling protocols have been proposed for the efficient control of multimedia communication services. However, the model and architecture of multimedia conferencing which is currently being standardized by IETF is not suitable to meet the requirement for the management of large-scale multimedia conferencing service. In this article, we have presented a management mechanism for distributed architecture of large scale multimedia conferencing service which is based on SIP. The high scalability is achieved by the coordinated distributed conference control and associated media processing without disruption of services. The SIP-based control mechanism for achieving the scalability has been designed in detail. Finally, the performance of the proposed architecture has been evaluated by simulation.

1. Introduction

The recent advance of broadband communication and computing technologies accelerates the proliferation of Internet telephony services. Internet telephony services can provide not only traditional voice services, but also multimedia communication services. It can provide high quality multimedia conferencing services using the Internet application protocols at relatively low cost. In addition, it can easily be integrated with other important application service such as remote education, messenger service, and home security service, and so on. This makes the demands for multimedia conferencing to be gradually increasing.

In response to the previously mentioned trends, the IETF is currently making the standard architecture and protocols for the multi-media conferencing service. The multimedia conferencing service takes advantage of IETF standard protocols for Internet telephony such as H.323 [1], session initiation protocol (SIP) [2], and media gateway control protocol (MGCP) [3] for call control and signaling. The ITU-T's H.323 defines the terminals, protocols and other components to provide multimedia communication on a packet network [4]. SIP is an application layer signaling protocol standardized by IETF which defines initiation, modification, and termination of multimedia communication sessions among users [2]. Among these, SIP is gradually gaining popularity due to its simplicity and flexibility for the control and management of multimedia conferencing service.

A conferencing management signaling protocol allows a user to initiate a multimedia conference, to request to join an ongoing conference, to invite another user to enter an ongoing conference, to voluntarily leave a conference, to remove a participant from a conference, and to obtain data transmission privileges [5]. The stream processing associated with mixing and encoding of different types of media has been studied for multimedia conferencing on a centralized server model [6]. A conferencing framework for multimedia conferencing management should provide

Proceedings of the First International Conference on Distributed Frameworks for Multimedia Applications (DFMA'05)
0-7695-2273-4/05 $ 20.00 IEEE

a high degree of flexibility and adaptability, the security mechanism, the integration with the existing management system, and a high scalability [7].

In order to achieve these goals, a great deal of research on the conferencing models and management mechanisms using SIP has been conducted and several models for multimedia conferencing have been proposed in IETF SIPPING WG. However, the conferencing models of IETF standards have limitations in scalability to be applicable to the large-scale multimedia conferencing service [8]. Most models are built on a single centralized conference server for the effective control of the multimedia conference, but the centralized conferencing server model is not suitable for a large-scale multimedia communication environment since it may have problems of triangular transmission, possible communication bottleneck due to the traffic concentration to the server, and large processing overload at the server.

In this article, we have presented a new highly scalable distributed architecture for the efficient management of large-scale multimedia conferencing service which is based on SIP. The high scalability is achieved by the coordinated distributed conferencing control and associated media processing. This overcomes the limitations of the conferencing models of IETF SIPPING WG. The distribution of both conferencing control and media processing functionalities enables the conferencing management operation such as conference initiation, invitation and joining, and leaving to be efficiently accomplished.

In Section 2, we present the conferencing architecture which are currently proposed in IETF SIPPING WG.. In Section 3, we design the scalable architecture for large scale multimedia conferencing management. In Section 4, the specific signaling mechanisms have been designed for the distributed conferencing management. In Section 5, we evaluate the performance by simulation, and the concluding remark finally follows in Section 6.

2. Conferencing Architecture Using SIP

Over the past few years, IETF SIPPING working group has been working on standards for the conference management and control using IETF session initiation protocol (SIP)[4][9][10][11]. In Figure 1, we show the general conferencing architecture which is currently being standardized by IETF SIPPING working group. The conferencing architecture consists of a conference server and participants. A focus is a SIP user agent which is responsible for the management of the conference using SIP signaling protocols. The focus handles the requests from participants by referring to the conferencing polices which are stored in the membership and media policy databases. The membership and media policy databases are managed by the policy server. The conference policy contains the rules that guide the decision-making process of the focus for the management of various conference requests from the participants. A mixer is responsible for handling the multimedia streams, and generating output streams which can be distributed to participants. A mixer can be located either in the focus or in the participant's user agent. In both cases, a mixer is controlled by the focus.

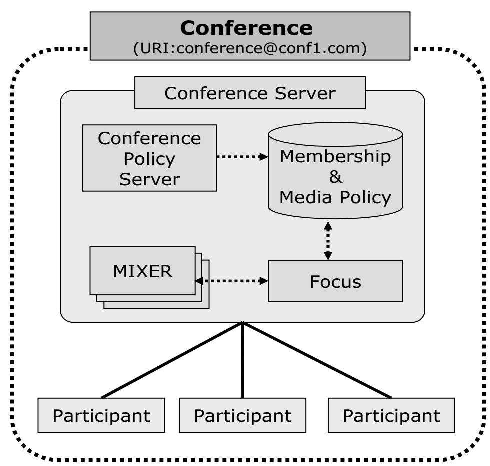


Figure 1. **The general conferencing architecture of tightly coupled conferencing model**

There are basically two multimedia conferencing models supported by SIP: a loosely coupled model and a tightly coupled model. In the loosely coupled model, each participant communicates to each other using IP multicast protocols. In the tightly coupled conferencing model, there is a centralized conference control server so that the conference can be managed more effectively. It supports a variety of conference control functions as well as media

Proceedings of the First International Conference on Distributed Frameworks for Multimedia Applications (DFMA'05)
0-7695-2273-4/05 $ 20.00 IEEE

mixing functions. Because of these characteristic of tightly coupled conferencing model, IETF pursues the tightly coupled conferencing model rather than the loosely coupled model. In the tightly coupled conferencing model, SIP signaling protocols are used for the control and management of the conference.

As shown in Figure 1, in the tightly coupled conferencing model, the conferencing system mainly consists of two parts: a conferencing server and participants. The tightly coupled conferencing model can be further classified into several models according to the locations of the focus and the mixer as follows: a centralized server model, an endpoint server model, a media server component model, a distributed mixing model, and a cascade mixers model [11].

In Table 1, we compare the characteristics of the various models. In a centralized server model, both the focus and the mixer are located together in a centralized conferencing server. In an endpoint server model, both the focus and the mixer are located together in one of participants. A participant plays the roles of both the server and the participant. In a media server component model, the focus and the mixer are separately located into two centralized conferencing servers. In a distributed mixer model, the focus is located in a centralized conferencing server and the mixer is located in participants. In a cascade mixer model, the focus is located in a centralized conferencing server, and the mixer is located in several distributed conferencing servers.

In the table 1, for the case of the cascade mixers model, although it is known to be designed for a large scale conference, the specific details on the control and signaling mechanisms has not been proposed. Furthermore, since the model depends on the centralized conference server, the communication bottleneck may occur due to the concentrated traffic to the single sever. When components fail in either server or mixer, it may be very difficult to recover the service due to the cascaded connection of the mixers. The distributed conference model proposed by the authors may eliminate all these problems by making the focus server to be configured in a distributed way, thereby achieving the greater scalability than that of the cascade model.

3. Scalable Distributed Architecture for Large Scale Multimedia Conferencing Management

We propose distributed conferencing architecture for the management of a multimedia conferencing service. In

Table 1. **The comparison of multimedia conferencing models**

	Location of Focus	Location of Mixer	Number of Servers	Relationship between Focus and Mixer	Scalability
Centralized Server Model	Central conferencing server	Central conferencing server	1	Co-located	Medium
Endpoint Server Model	One of participants	One of participants	0	Co-located	Small
Media Server Component Model	Central conferencing server	Central conferencing server	2	Separated	Medium
Distributed Mixing Model	Central conference server	Every participant	1	Separated	Medium
Cascade Mixers Model	Central conferencing server	Distributed conferencing server	Many	Separated	Large
Distributed Conferencing Server Model (Proposed Model)	Distributed conferencing server	Distributed conferencing server	Many	Separated	Very large

Proceedings of the First International Conference on Distributed Frameworks for Multimedia Applications (DFMA'05)
0-7695-2273-4/05 $ 20.00 IEEE

Figure 2, we describe the distributed conferencing architecture which vertically consists of three tiers: a conference management tier, a mixer tier for multimedia stream processing, and the participants. The salient feature of the architecture is that the conference management tier is configured in a distributed way. It is obtained by the extension of the tightly coupled conference server model in Figure 1.

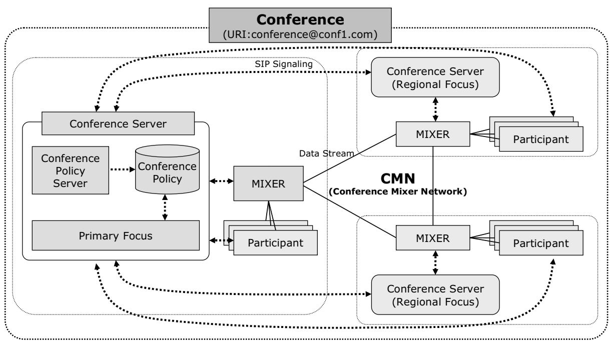


Figure 2. **A distributed conferencing architecture**

In the distributed conferencing architecture, a conference consists of several local conference servers and mixers. Each local conference server contains a regional focus which is responsible for the management of the corresponding local conferencing service. The regional focus also manages the corresponding mixers in the region for load sharing and media streaming. In the architecture, one of the regional focuses is designated as a primary focus. The conference is horizontally comprised of one primary focus, several regional focuses for signaling and controlling of the conference, and several separated conference mixers for multimedia stream treatment. The conference mixer handles mixing and redistribution of multimedia streams such as conference video and audio streams. The set of mixers involved in the conference constitute a network called as a conference mixer network (CMN).

The primary focus is responsible for the control of the whole conference in an integrated way. It sets up the CMN and modifies the CMN. It also controls the access to the conference server so that the participants should first get the permission from the primary focus to participate in a specific conference session. The primary focus announces the conference session information using session announcement protocol (SAP) [12], and handles participation requests. The primary focus can add and delete mixers according to the scale of conference, so that it can compose the CMN properly. Thus, the control and mixing operations are distributed in the proposed distributed conferencing architecture, so that the processing overload and traffic concentration can be reduced. These features can greatly enhance the scalability of the conferencing system.

Since the CMN can be configured independently of participants, participants don't have to take care of the composition of the CMN. This makes the signaling and stream procedure of the centralized conferencing model to be used without modification. Each participant can obtain adjacent conference server (ACS) information using service location protocol (SLP) [13], and the mixer of each participant can encode and transmit data. Additionally, the triangular transmission which is caused by accessing the remote server can be eliminated, and delay and traffic in the core network can be reduced accordingly.

4. Management Mechanism for Distributed Multimedia conferencing

The distributed conferencing model relies on the SIP signaling methods and its extensions to manage a conference and the CMN. A user can join a conference by submitting a connection request to the primary focus.

Figure 3 shows the procedure of conference invitation call flow using SIP signaling methods. When an invitee logs on, the invitee sends to the primary focus an INVITE message which contains the ACS information. Since the ACS information contains the local conference server which is adjacent to the invitee, the primary focus identifies whether the corresponding ACS mixer is participated or not. If the ACS mixer associated with the invitee is participating in the conference, the primary focus allows, using the REFER message [14], the invitee to be connected to that mixer.

In the case where the invitee's ACS mixer does not participate in, the primary focus tests whether an additional mixer is necessary or not. This process is done by using the *minimum mixer participant* value which is defined by the membership policy. If the number of

Proceedings of the First International Conference on Distributed Frameworks for Multimedia Applications (DFMA'05)
0-7695-2273-4/05 $ 20.00 IEEE

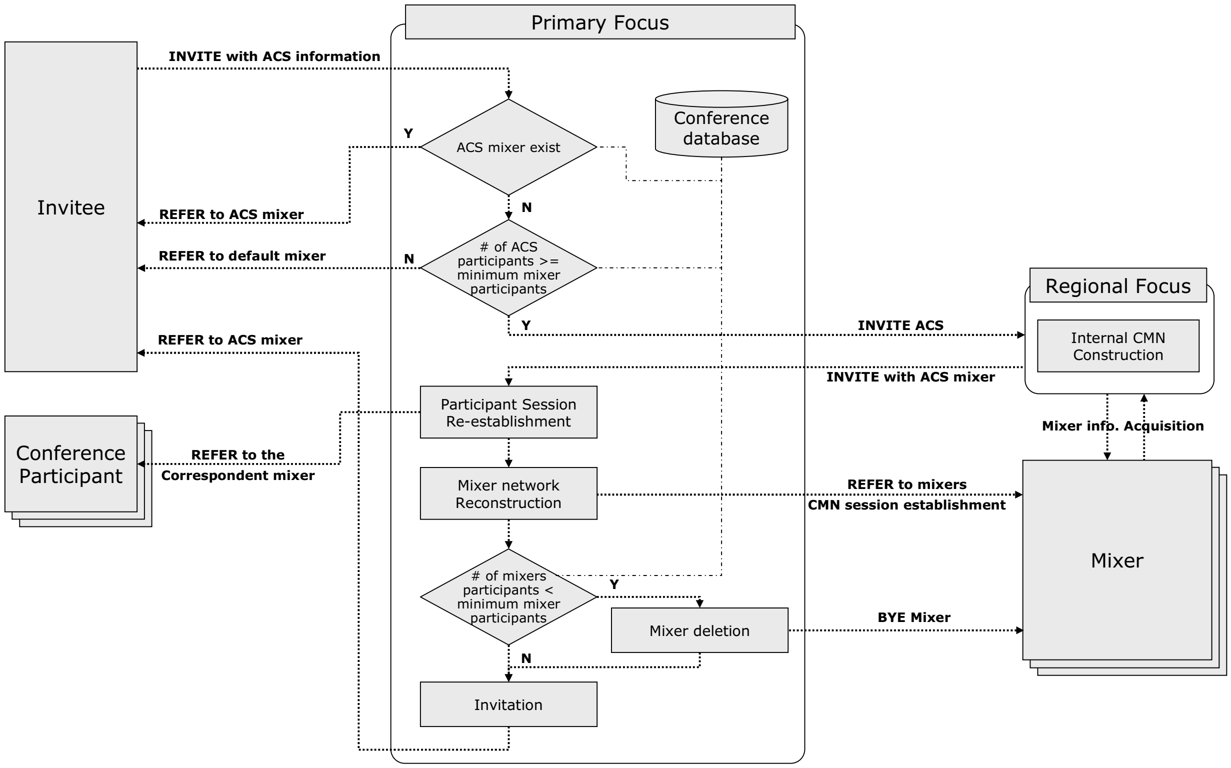


Figure 3. **A Signaling procedures for distributed conferencing management**

participants having the same ACS information is less than the *minimum mixer participant* value, the primary focus makes the invitee to be connected with the default mixer. And if the number of participants having the same ACS information is greater than the *minimum mixer participant* value, the primary focus requests the regional focus to make a new mixer to be available. After a new mixer is getting involved in the conference, the primary focus redistributes the participants to the relevant mixers and changes the configuration of the CMN. When the participant redistribution process ends, the primary focus identifies the number of participants which are connected with each mixer. If the number of participants is less than *the minimum mixer participant* value, the primary focus deletes that mixer, and re-connects the participants at the deleted mixer to other proper mixers. After all these procedures are performed, the primary focus then requests the invitee to be connected to a new participating mixer.

We describe below the detailed signaling procedures for the conference management in the proposed distributed conferencing model.

- **Conference Initiation**

In the case of creating a new conference, a participant requests the regional focus to add a mixer using INVITE message, and then participates in a conference by session re-establishment to the mixer. At this time, the regional focus will be a primary focus until the conference is over. Figure 4 shows the signaling procedure of conference initiation in which the participants A and B initiate a new conference.

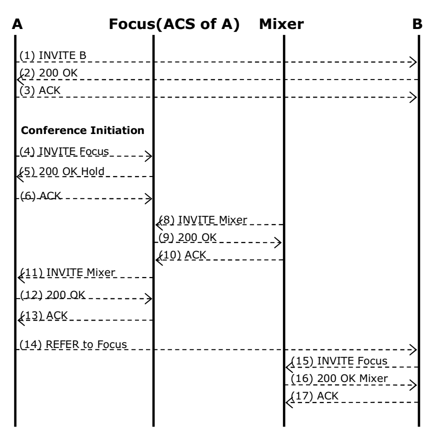


Figure 4. **Signaling procedures for the initiation**

Proceedings of the First International Conference on Distributed Frameworks for Multimedia Applications (DFMA'05)
0-7695-2273-4/05 $ 20.00 IEEE

- **Invitation and Joining**

After the conference is initiated, If a new mixer is required, the primary focus sends an INVTE message to the new mixer. After the new mixer is allocated in the conference, the CMN is re-configured accordingly, and the primary focus may re-assign the conference participants to a new mixer, which can be determined by the membership policy. In order to do that, the primary focus sends a REFER message to the relevant participants to re-assign the participants to the new mixer. During the re-assignment operation, the re-assigned participant should terminate the session with the old mixer using BYE message.

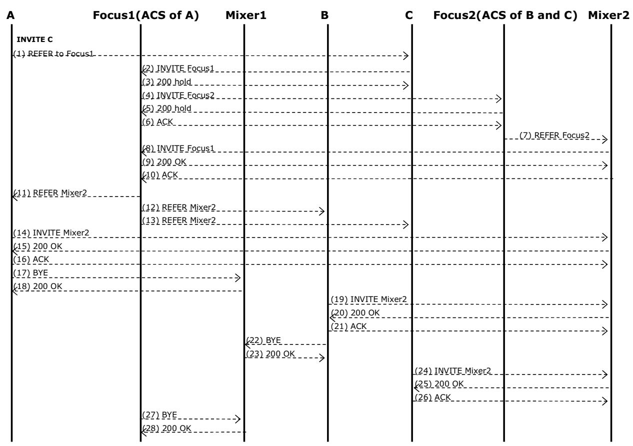


Figure 5. **Signaling procedures for invitation**

Figure 5 shows an example of these signaling procedures for the management of invitation and joining in which the participant A invites a new participant C. It is assumed that the participants A and B are already in the conference session after initialization.

- **Leaving**

After the participant leaves the conference, if the number of participants assigned to a mixer is less than the *minimum mixer participant* value, the mixer which has been associated with the left participant should also be disconnected from the conference. This requires that the remaining participants associated with the disconnected mixer should be re-assigned to another mixer.

The primary focus is responsible of this re-assignment, and it sends a REFER message to a new mixer. The remaining participants can then keep conferencing by session re-establishment to a new mixer. After all the remaining participants are reconnected with the new mixer, the primary focus sends a BYE message to the old mixer, and the old mixer can leave the conference. Figure 6 shows the signaling procedures for leaving and CMN reconfiguration in which the participants C and B leave the conference.

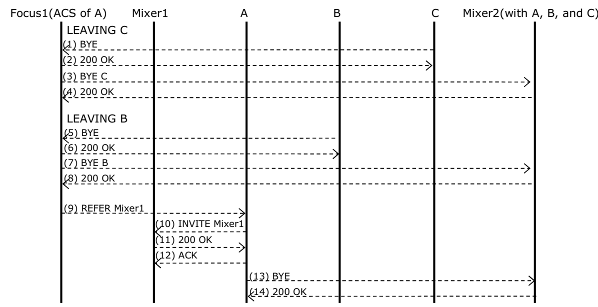
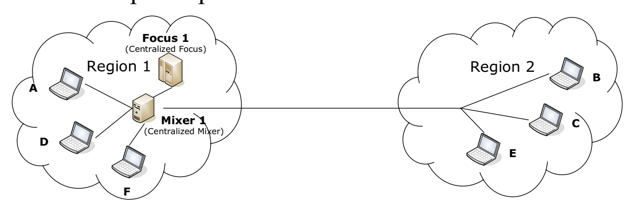


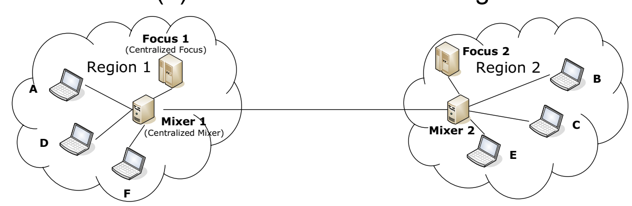
Figure 6. **Signaling procedures for leave and CMN reconfiguration**

5. Performance Evaluation

In the global conferencing environment, the participants are located in different regions: Region 1 and Region 2. The nodes A, B, C, D, E, and F represent the conference participants.



(a) Centralized conferencing



(b) Distributed conferencing

Figure 7. **Test networks for performance evaluation**

In the centralized mode of conferencing management, the Focus1 and the Mixer 1 play the role of the centralized server as shown in Figure 7 (a), whereas in a distributed model, the management operations are performed by the

Proceedings of the First International Conference on Distributed Frameworks for Multimedia Applications (DFMA'05)
0-7695-2273-4/05 $ 20.00 IEEE

primary focus, Focus 1, and the stream distribution is performed by Mixer 1, and Mixer 2 in a distributed way as shown in Figure 7 (b). The ACS of the nodes A, D, F is Focus 1, and that of the nodes B, C, E is Focus 2, respectively. The network parameters for simulation are described in Table 2.

Table 2. Simulation environments

Network Parameter for simulation	Delay
Region-to-region transmission delay	50ms
Transmission delay in the same region	10ms
Focus-to-mixer signal transfer delay	5ms

Figure 8 shows the delay characteristics when the participant A invites other participants B, C, D, E, and F. The delay is measured in the average signaling completion time. Initially, the participant A invites the participant B, and in this case, the average signaling delay is identical in both distributed and centralized models. This is indicated by the delay due to inviting the participant B. However, when the participant C is invited, the mixer network configuration processing is needed. A new mixer Mixer2 is created to handle the media requests from the new participant C, and the existing participant B should be assigned to the Mixer2. This re-adjustment generates some delay so that the invitation completion time for the participant C is large as shown in Figure 8. This pattern of re-configuration continues to invite other participants D, E, and F, and the related signaling delay times are shown in Figure 8.

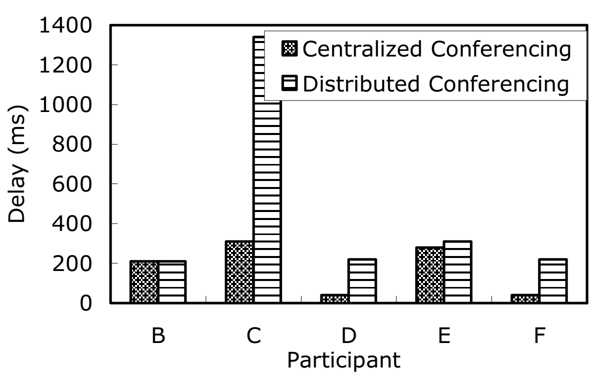


Figure 8. **Average signaling delay for invitation**

As shown in Figure 8, the distributed conferencing model creates larger delay time than that of the centralized conferencing model. However, when the number of participants is large, and they are grouped and located in different regions, the signaling delay can be reduced since the region focus can perform the conferencing management functions.

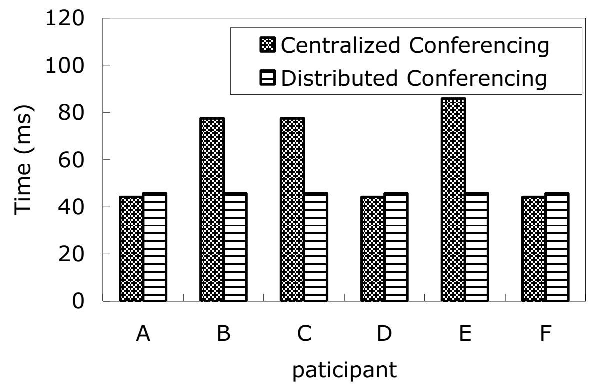


Figure 9. **Average stream transmission delay**

Figure 9 shows the average delay for broadcasting multimedia stream among participants. In case of A, D, and F, the average delay of the centralized model is a little less than those of the distributed conferencing model. This is due to the processing delay at the Mixers2 to deliver the streams to the participants in the region. However, for the cases of B, C, and E, the delay of the distributed model is smaller than that of the centralized model. This is because the multimedia stream can be locally transferred in the distributed model without passing through the centralized mix. For example, the multimedia stream from the participant B can be directly transferred to the participants C and E through the Mixer2. This reduces the average delay for the transmission of multimedia streams. In a distributed conferencing model, the signaling mechanism is more complex than that of the centralized conferencing model. However, the stream transmission delay of the distributed conferencing model is generally smaller than that of the centralized conferencing model.

Figure 10 shows the result of measuring the processing load for encoding/decoding and mixing at the mixer for the transmission of multimedia stream traffic. Both the processing load of the centralized conferencing model and that of the distributed conferencing model are shown in comparison. As shown in Figure 10, the

Proceedings of the First International Conference on Distributed Frameworks for Multimedia Applications (DFMA'05)
0-7695-2273-4/05 $ 20.00 IEEE

processing load of the centralized model drastically increases as the number of participants increase, while in the distributed model, the processing load is almost constant. This illustrates the fact that the distributed model performs better than the centralized model with regard to the scalability.

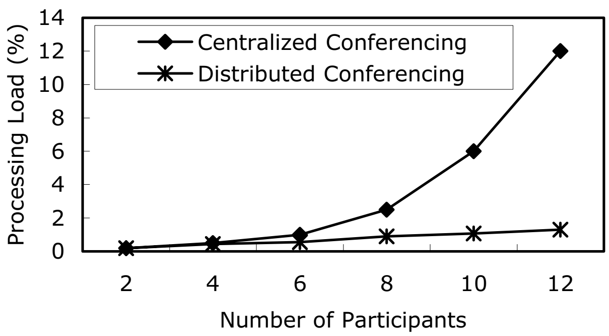


Figure 10. **Average processing load of mixer**

6. Conclusion

In this paper, we have presented a management mechanism for distributed architecture of multimedia conferencing service. We have first studies the features of current IETF standards for the multimedia conferencing service using SIP, and investigate the limitation of the standards to be applied for the multimedia conferencing service. We have then designed the scalable distributed conferencing architecture for the efficient control and management of the multimedia conferencing service. We specifically design signaling procedures conferencing mechanisms by extending the current IETF standards for the management of conferencing service. In order to show the efficiency of the proposed architecture, we have evaluated the performance by simulation. The simulation results shows that the proposed distributed architecture performs better in delay time and processing load distribution than those of the centralized conferencing model.

The scalability is achieved by the distributed conference control and mixer networks which reduce the processing overhead at the conference servers and eliminate the communication bottleneck due to the media mixing and transmission. The proposed distributed conferencing architecture can be applied not only for the multimedia conferencing service but also for various other multimedia services such as remote education and distributed on-line games. Further work may include the research on the multimedia distributed conferencing model in wireless Internet environment.

References

[1] ITU-T Rec. H.323. “Packet based Multimedia Communications System”, v2, 1998.

[2] J. Rosenberg, H. Schulzrinne, G. Camarillo, A. Johnston, J. Peterson, R. Sparks, M. Handley, and E. Schooler, "SIP: Session Initiation Protocol", RFC 3261, June 2002.

[3] F. Andreasen and B. Foster, “Media Gateway Control Protocol(MGCP) version 1.0”, RFC 3435, January 2003.

[4] ITU-T “Recommendation H.323: Packet-based Multimedia Communications Systems”, ITU, Geneva, November 2000.

[5] C. Ziegler and G. Weiss, “Multimedia Conferencing on Local Area Networks”, *IEEE Computer*, vol. 23, no. 9, pp. 52-61, September 1990.

[6] K. Singh, G. Nair and H. Schulzrinne, "Centralized Conferencing using SIP", Proceedings of the 2nd IP-Telephony Workshop, April 2001

[7] A. Keller and H. Reiser, “Dynamic Management of Internet Telephony Service: A case Study based on JavaBeans and JDMK”, Proceeding of Enterprise Distribute Object Computing Conference, 1999.

[8] J. Rosenberg and H. Schulzrine, “Models for Multimedia Conferencing in SIP”, Internet Draft, IETF, Nov. 2000.

[9] O. Levin and R. Even, “High Level Requirements for Tightly Coupled SIP Conferencing”, Internet-Draft, SIPPING WG, April 2003

[10] J. Rosenberg, “A Framework for Conferencing with the Session Initiation Protocol”, Internet-Draft, SIPPING WG, May 2003

[11] A. Johnston and O. Levin, “Session Initiation Protocol Call Control - Conferencing for User Agents,” Internet-Draft, SIPPING WG, June 2003

[12] M. Handley, C. Perkins and E. Whelan, “Service Announcement Protocol,” RFC 2974, October 2000

[13] J. Veizades, E. Guttman, C. Perkins and S. Kaplan, “Service Location Protocol,” RFC 2165, June 1997

[14] R. Sparks, “The Session Initiation Protocol (SIP) Refer Method,” RFC 3515, April 2003.

Proceedings of the First International Conference on Distributed Frameworks for Multimedia Applications (DFMA’05)
0-7695-2273-4/05 $ 20.00 IEEE