# A Loosely Coupled Video Conference Based on SIP

T.-L. Kao[*1], Tzung-Jung Tsai[#2] , K.-T. Wu[#3]

[*]*Communication Engineering Department, I-Shou University*
*No. 1, Sec. 1, Syuecheng Rd., Dashu Township, 840, Kaohsiung County, Taiwan, R.O.C.*
[1]tlkao@isu.edu.tw

[#] *Electrical Engineering Department, National University of Kaohsiung*
*Kaohsiung University Rd., Nanzih District, 811, Kaohsiung, Taiwan, R.O.C.*
[2]M0975126@mail.nuk.edu.tw
[3]ktwu@nuk.edu.tw

*Abstract*— **The improvement in Internet bandwidth and the progress in computing technology have accelerated the maturity of video conference system. Since the cost of MCU (Multipoint Control Unit) is expensive, the video conference system based on software is adapted for cost considerations. There are many conference system architectures based on SIP but most of them are centralized. Although these architectures are fine from the standpoint of operations, there are some drawbacks in these centralized architectures such as scalability and robustness. The bottleneck may occur due to the fact that signalling and streaming traffic are directed to the centralized server and may be congested in the server. In this paper, we propose an architecture for which control signalling and multimedia streaming are processed separately, that is, centralized in control signalling and distributed in the multimedia streams.**

*Keywords* — **SIP, Conferencing, Video Conference System**

## 1. INTRODUCTION

The improvements in the Internet bandwidth and the progression in computing technology have accelerated the maturity of multimedia applications on the Internet, e.g. IPTV, VoIP etc. Although VoIP (Voice over Internet Protocol) have been developed many years. There are some advantages of VoIP such as reducing communication and infrastructure costs, provide services that may be more difficult to implement by using PSTN. The video conference system is one of these multimedia applications.

Many international business companies are established in different countries. These companies require the reduction of costs and labour in order to run branch offices in lots of areas or countries. However, it will raise the issues of the communication among the headquarters and the branch offices, and the cost of communication is sometimes a big expenditure in the operations in these companies. Therefore the video conference system will be the first choice to reduce costs for business travels. The video conference system provides a mean for conference attendees to see and talk with each other even though they are locating in different places.

Presently the main protocols of VoIP are SIP (Session Initiation Protocol) RFC-3261 of IETF (Internet Engineering Task Force) and H.323 of ITU-T (International Telecommunication Union – Telecommunication Standardization) [1] [2] [7]. The H.323 specifies that the video conference function required MCU (Multipoint Control Unit). This architecture is complicate, the cost of MCU is expensive, and the flexibly is low. In contrast, SIP specifies a Framework for Conferencing Session Initiation Protocol (RFC 4353) [3], which is more flexible and less complicate. It uses the original benefits of SIP to develop the video conferencing solutions. There are two models in RFC 4353, i.e., Tightly Couple Model (Centralized) and Loosely Coupled Model (Distributed). The Tightly Coupled Model is fine for deployment, yet it is a centralized architecture [4]-[6] with less scalability and robustness. A flexible architecture is proposed in this paper for which the control signalling is centralized and multimedia streams are distributed.

This paper is structured as follows. Section 2 presents the background and the related work of conference framework. Section 3 presents the architecture of the proposed system, while Section 4 describes the implementation. And test results in Section 5 are described. Finally, we summarize our results in Section 6 and outline the items for future work in Section 7.

## 2. BACKGROUND

SIP has been defined for the establishment, maintenance, and termination of connections between one or more endpoints. In RFC 4353, there are two model of conference based on SIP: Tightly Coupled Model and Loosely Coupled Model.

A loosely coupled model is designed to minimize the negotiated signalling amongst these participants. Loosely coupled model usually uses the IP multicast for distribution of conference memberships. In the tightly coupled model, there will be a single user agent, referred to as a focus, as the role of the centralized manager of the conference, so that the conference can be managed more effectively. Even though this architecture has its advantages such as a low latency in connection establishment, easy to develop and deploy. It has some vital drawbacks such as being subject to DoS (Denial of Service) attacks, lack of robustness and scalability, etc. However, the tightly coupled model is adapted rather than loosely coupled.

As shown in Fig. 1, the conferencing framework with SIP is specified in RFC 4353 by IETF. It consists of two parts: a conference server and participants.

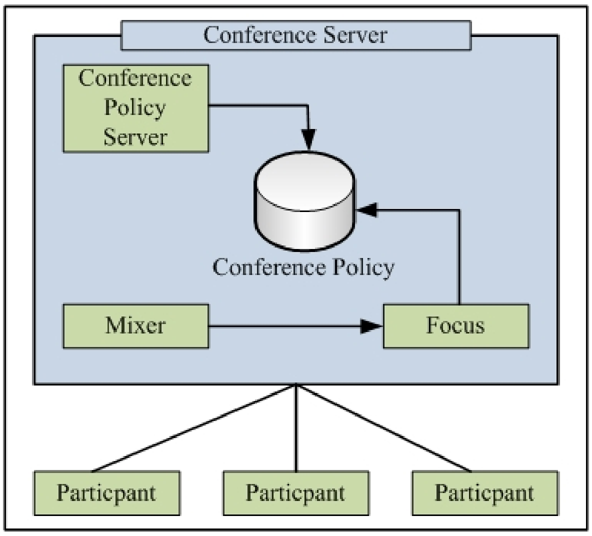


Fig. 1 RFC 4353 conferencing architecture

### 2.1. Logical Elements

The RFC 4353 defines logical elements that a conference server should have. There are **Focus**, **Mixer**, **Conference Policy Server**, etc.

- **Focus:** The focus is a SIP user agent that is responsible for negotiating with each participant by using SIP signalling in the conference. It handles the requests from the participants by referring to the conference policies which are stored in the membership and media policy database.

- **Mixer:** A mixer receives a set of media streams of the same type, and combines these media in a type-specific manner, redistributing the resultant media to each participant.

- **Conference Policy:** Conference policy is a complete set of rules governing a particular conference, such as the white list, black list, and the maximum members in a conference room.

- **Conference Policy Server:** A conference policy server is a logical function that can store and manipulate the conference policy. This logical function is not specific to SIP. It refers to the component that interfaces a protocol to the conference policy.

- Participant: The software element that connects users or automata to a conference. It implements a SIP user agent, but may also implement non-SIP-specific mechanisms for additional functionality.

### 2.2. Models of Video Conferencing

The tightly coupled conferencing model can be further classified into several models according to the locations of the focus and the mixer as follows: (1) centralized server model, (2) endpoint server model, (3) media server component model, (4) distributed mixing model and (5) cascade mixers model.

In Table 1, a comparison of video conferencing models is tabulated. In the centralized server model, both of the focus and the mixer are placed in a centralized conferencing server. In endpoint server model, both of the focus and the mixer are placed at location of the participant. In media server component model, the focus and the mixer are placed in two different conferencing servers separately. In distributed mixers model, the focus is allocated in a centralized conferencing server, while the mixer is allocated to each participant. In a cascade mixer model, the focus is placed in a centralized conferencing server, and the mixers are allocated to multiple distributed conferencing servers.

**TABLE 1**
**COMPARISON OF VIDEO CONFERENCING MODELS**

| | Location of Focus | Location of Mixer | Number of Servers | Relationship between Focus and Mixer | Scalability |
|---|---|---|---|---|---|
| **Centralized Server Model** | Centralized conferencing server | Centralized conferencing server | 1 | Co-located | Medium |
| **Endpoint Server Model** | One of participants | One of participants | 0 | Co-located | Small |
| **Media Server Component Model** | Centralized conferencing server | Centralized conferencing server | 2 | Separated | Medium |
| **Distributed Mixing Model** | Centralized conferencing server | Every participant | 1 | Separated | Medium |
| **Cascade Mixers Model** | Centralized conferencing server | Distributed conferencing server | Many | Separated | Large |
| **Proposed Model** | Centralized conferencing server | None | 1 | N/A | Medium |

Even though the cascade mixer model is designed for a large scale conference, it depends on a centralized conference server and many distributed conferencing servers, a communication resources squander may occur due to the fact that too many servers have been exercised [12].

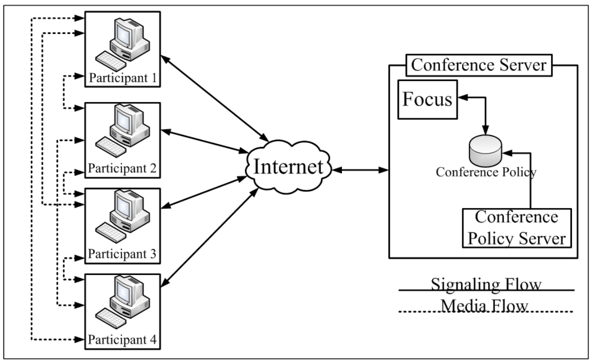


Fig. 2 Proposed architecture of Video conference system

# 3. THE ARCHITECTURE OF SYSTEM

According to RFC 4353 conferencing framework, some changes are proposed in its original architecture. The original centralized signalling management, which can manage the server effectively, is modified to avoid the communication bottleneck due to the concentration of media streaming traffic to the signal server which may bring about system's defects such as robustness and scalability. As shown in Fig. 2, we propose an architecture of multimedia video conferencing, which provides a distributed method to transmit the media streaming and a centralized approach to manage the signalling negotiation.

The whole architecture consists of Participant UA and Conference Server.

## 3.1. Participant UA

According to SIP, the UA (User Agent) is the user endpoint in the SIP environment. It uses request and responds to initiate and terminate a dialogue. The UA can be a SIP phone or a softphone in the personal computer. It composes of UAC (User Agent Client) and UAS (User Agent Server) logically. The UAC is responsible for establishing requests and the UAS is responsible for responding the requests that from UAC.

The ABTO (Advanced Brainstorming & Technology Outsourcing) SIP VoIP SDK (Software Development Kit) is used to develop the proposed system [10]. A SIP UA program is designed to fit in with our architecture. Users can conference with the personal computer and webcam. When the program is started, users need to process authentication and register to a conference server. After that, users can get a user list of all of the participants. Each user can invite others to join the conference. The main purpose of using ABTO's SIP VoIP SDK is to build a SIP UA that specified by SIP so that our program can communicate with the conference server. There is an example of participant UA#1, #2, #3 registers to the conference server as showing in Fig. 3.

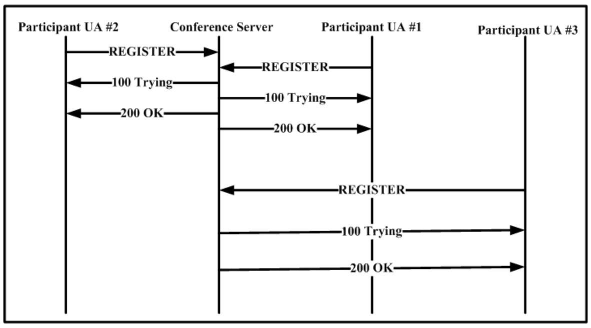


Fig. 3 Participant UA connect to conference server

## 3.2. Conference Server

The conference server consists of: Focus, Conference Policy, and Conference Policy Server. The focus is developed by Open Source Asterisk [8], it includes the UA which is the conference server to communicate with users. Conference policy is designed by MySQL [9]. MySQL is a database, responsible for storing conference policy and the information of participants.

However, the conference policy can use the white list or block list to manage the conference. The Asterisk is an Open source PBX (Private Branch eXchange) system. It is responsible for providing SIP voice call or video service as a normal SIP server to complete the whole VoIP telecommunication system.

# 4. IMPLEMENTATION

## 4.1 Developing Environment

(1) Operation System: Windows XP, CentOS
(2) Build Tool: Visual Studio 2008
(3) Database: MySQL 5.0.43
(4) Focus: Asterisk 1.4.26.2

## 4.2 Connection Function of Participant UA with Conference Server

As shown in Fig.4, the Participant UA program is initiated to connect to the conference server. This action includes the register and authentication. When these actions are complete, the information of participants will be updated at the conference server. Finally every participant can access this information from the conference server.

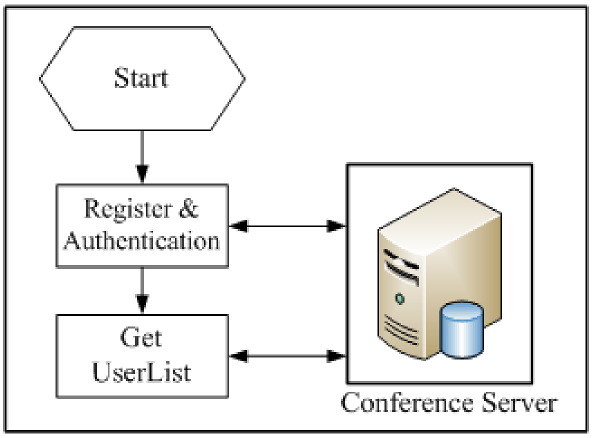


Fig. 4 Flow chart of the connection function

## 4.3 Conferencing function of Participant UA

The flow of the program is shown in Fig. 5. It primarily includes four parts: video processing, audio processing, client function, server function.

- Video & Audio processing: In this part, Windows API is used to obtain the webcam device. And then storing these data into a memory stream for sending to other participants.
- Client & Server function: This function is a network program. It not only opens the socket on participant's computer, but also initiates the TcpListener to listen on the port number continuously [11].

The transformation from IP address into username is a mapping method, which uses the data base from MySQL.

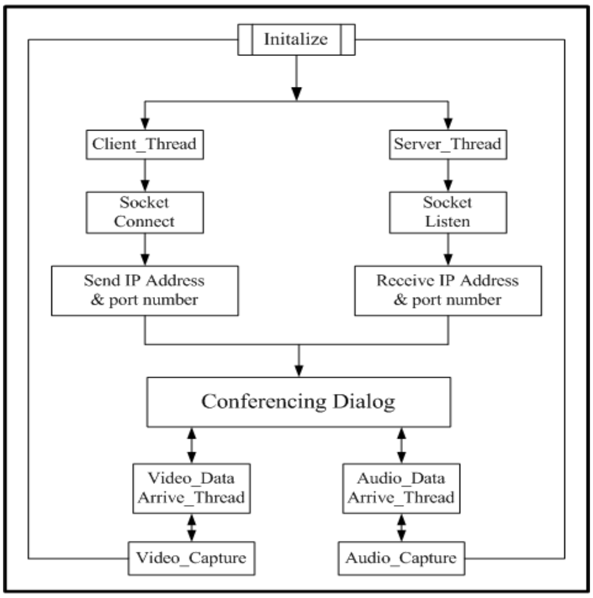


Fig. 5 The flow of the Participant UA

# 5. TEST RESULTS

The architecture of the testing environment is shown in Fig. 6. There are three participants located at same local area network behind a NAT (Network Address Translation) device. Meanwhile the conference server is placed on the public network. There is a picture from one of the participants' computer screens shot as shown in Fig. 7. This participant's username is 55688. His video is on the left side of Fig. 7, the middle video of Fig. 7 is the user 41001. And the video of the right side of Fig.7 is the user 41002. Every participant has the right to decide to release his video to other participants.

The lower left corner of Fig. 7 is the block of SIP UA for register and authentication. The user must input his identification, such as account, password. Then the user can connect to conference server to do registration and authentication. After that action, each user can press the Get User List button to get the information of every participant. This list includes username and IP address.

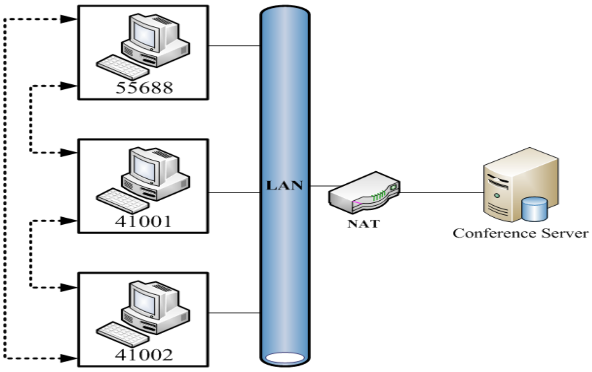


Fig. 6 Architecture of proposed testing environment


Fig. 7 The screen shot of one of the participants

## 6. Conclusions

Although the conferencing functions based on SIP have been developed several years ago, most of the architecture is tightly coupled model. There are some drawbacks such as scalability, robustness, and DoS attacks in this model. In this paper, we propose an architecture to minimize these drawbacks in a tightly coupled model. It takes the advantage of powerful computing technology and the broad bandwidth of internet. It is used for point to point transmission way to achieve the conferencing function and solve the jammed media streaming in the centralized conference server.

## 7. Future work

There are some issues in this paper, such as the mechanism to manage conference policy, the other multimedia streaming transmission, and distributed focus deployment. The prototype of this architecture is implemented and testing will be done after the completeness of the design.

## References


[1] J. Rosenberg, H. Schulzrinne, G. Camarillo, A. Johnston, J. Peterson, R. Sparks, M. Handley & E. Schooler. *SIP: Session Initiation Protocol. RFC 3261*, 2002.

[2] J. Rosenberg & H. Schulzrine. *Models for multimedia conferencing in SIP. Internet Draft*, IETF, 2000.

[3] J. Rosenberg, *A Framework for Conferencing with the Session Initiation Protocol (SIP). RFC 4353*, IETF, 2006.

[4] NI Jian-yun & LUO Jing, *Design of multimedia conference control system based on SIP*, Journal of Communication and Computer, Vol.4, No.3, pp.55-60, 2007.

[5] R Venkatesha Prasad, Richard Hurni & H S Jamadagni. *A scalable distributed VoIP conferencing using SIP*. Proceedings of the Eighth IEEE International Symposium on Computers and Communication, 2003.

[6] Ying Jie & Shen Sun, *Implementation of Multipoint Video Conference in Software*, Circuits and Systems, 2006. APCCAS 2006. IEEE Asia Pacific Conference, pp.1980-1983, Dec. 2006.

[7] *H.323, Packet-based multimedia communications systems*, ITU-T Standards, Nov. 2000.

[8] Asterisk, http://www.asterisk.org/.

[9] MySQL, http://www.mysql.com/.

[10] ABTO (Advanced Brainstorming & Technology Outsourcing), http://www.abtollc.com.

[11] MSDN (Microsoft Developer Network), http://msdn.microsoft.com/zh-tw/netframework/default.aspx.

[12] 許晃瑋、盧俊佑、蔡孟倩、王楓宜、蘇暉凱, *階層式 SIP 網路多媒體會議嵌入式系統之設計與實作*, National Symposium on Telecommunications 2008 (NST2008), 2008.