SIP Based Intelligent Multimedia Conferencing

Yogesh Pingle¹ and Arjun Jaiswal²
¹Information Technology Department, Vidyavardhini’s College of Engineering and Technology
²Computer Engineering Department, Rizvi College of Engineering
¹yogesh.pingle@vcet.edu.in and ²arjun.jaiswal1985@gmail.com

ABSTRACT
Multimedia conferencing, or video conferencing, is an important application of real time media. In video conferencing, people at different sites are brought together for a meeting by transmitting real time audio, video and collaboration data on communication channels. Video conferencing are widely used in telecommuting, distant collaboration, distant learning, career services etc. With the fast development of technology, video conferencing will have broader impact and huge potential market.

As an advanced communication tool, inter-connectivity and inter-operability require that the video conferencing devices involved in the same conference can talk to each other, i.e., they comply with some common standard. Video conferencing standard is an umbrella set of standards because it not only has to specify audio and video coding standards, but also needs to address call control, conference management, media packetization and delivery.

There are two major categories of video conferencing standards, H.32x series from the telecommunication world, standardized by ITU (International Telecommunications Union), and SIP (Session Initial Protocol) based video conferencing standard from the Internet world, recommended by IETF (Internet Engineering Task Force).

KEYWORDS
H.32x, SIP, Multimedia Conferencing

1. INTRODUCTION
Multimedia conferencing is rapidly gaining interest in the field of communication. There are already a few products of multimedia conferencing based on H.32x standard. SIP, which is a more feasible protocol, is put on the agenda of being the call signaling protocol for conferencing. Most of the researches on SIP based conferencing, however, have still remained on theories.

Multimedia Conferencing is advantageous to people who don’t want to spend their time and money flying all over the world for face-to-face meetings. International standardization bodies have defined protocols for multimedia conferencing. On comparing SIP and H.32x protocols, we found that H.32x is widely deployed only because of its early adoption by the market, but has several problems. On the contrary, SIP is more lightweight, flexible and extensible. It is a text-based protocol which can easily interact with other internet protocols [1].

So, it is intended to implement a multimedia conferencing system using SIP. Once the implementation of conferencing system is complete, enough information will be available regarding the type of conferencing that a user does and with which user does he/she interacts most of the time. Hence the system is named as SIP Based Intelligent Multimedia Conferencing.

2. THEORETICAL ISSUES
Handling conferencing data packets is always a problem, as participants work with different platform desktops. Also, transmitting media data across the network in real-time requires high throughput. It’s better to compensate for data loss than to compensate for large delays. TCP is a transport-layer protocol designed for reliable data communications. When a packet is lost or corrupted, it’s retransmitted. The overhead of guaranteeing reliable data transfer slows the overall transmission rate. For this reason, underlying protocols other than TCP (e.g. UDP) are typically used for streaming media [2].

SIP runs on both TCP and UDP, and in fact, can be extended to run on almost any transport protocol. H.32x only runs on TCP, which loads the servers. So, the basic idea behind implementing this system is to overcome the disadvantages of current systems that use H.32x protocol for multimedia conferencing [6].

3. SCOPE
Transmitting media data across the network in real-time requires high throughput. It’s better to compensate for data loss than to compensate for large delays. One major drawback of the H.32x protocol is its lack of scalability. When extended to world-wide networks, SIP has many advantages [3] [4]:

- **Loop detection**: When trying to locate a user over several domains, loops can occur. H.32x has no support for loop detection. Loops are easily detected using SIP headers, as they specify all proxies that have handled the SIP packet.
- **Distributed control**: In H.32x, there is a need for a central point when performing multi-user calls, which means that this central point must be dimensioned for the size of the call. SIP sessions are distributed, making the need of this central point disappear.
- **Small connection overhead**: Establishing a connection using H.32x takes about three times the data and turnarounds compared to when using SIP. TCP is a transport-layer protocol designed for reliable data communications. When a packet is lost or corrupted, it’s retransmitted. The overhead of guaranteeing reliable data transfer slows the overall transmission rate. For this reason, underlying protocols other than TCP (e.g., UDP) are typically used for streaming media. SIP runs on both TCP and UDP, and in fact, can be extended to run on almost any transport protocol. H.32x only runs on TCP, which loads the servers. So, the basic idea behind implementing this system is to overcome the disadvantages of current systems that use H.32x protocol for multimedia conferencing [4].

4. **H.32x STANDARDS FOR CONFERENCING**

Handling conferencing data packets is always a problem, as participants work with different platform desktops. This calls for a standard to which the software must conform. To facilitate the compatibility among various vendor solutions, ITU developed H.320 standard series, including H.322, H.323 and H.324. The core functions of H.32x series are as the following [4]:

- H.322 - applications over LANs that provide a guaranteed bandwidth
- H.323 - applications over LANs that provide a non-guaranteed bandwidth
- H.324 - applications over the PSTN and mobile phones

5. **SIP CONFERENCING**

The goal is to design and implement Video Conferencing over the network. The software will not only guarantee better real time service, but also good performance with respect to the playback of the streamed media data. For this, a way will be created to package MPEG video into RTP packets. The SIP conferencing architecture as shown below in Figure 1 consists of a conference server and a participant. A focus (SIP user agent) is responsible for management of conference using SIP signaling protocol. 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 notification service is a logical function provided by the focus. The focus can act as a notifier, accepting subscriptions to the conference state, and notifying subscribers about changes to that state. A mixer is responsible for handling the multimedia streams, and generating output streams which can be distributed to participants. It is also controlled by the focus. Because of the strongpoint described about SIP, surely SIP will be selected as a call signaling protocol.

The SIPPING conferencing framework uses SIP protocol as call signaling for session management and control means for implementation. Call signaling is a process that is used to set up a connection in a network. It includes: authentication, authorization and communications security (to help prevent eavesdropping or manipulation of message content). The highlight of this design is conference policy, which contains the rules that guide the decision-making process of the focus for the management of various conference requests from the participants [1].

6. **BASIC REAL TIME TRANSPORT PROTOCOL**

RTP is a transport protocol for real-time applications. This protocol is used for the transport of real-time data, including audio and video. RTP provides support for applications with real-time properties such as continuous media (e.g., audio and video), including loss detection, security and content identification. RTP does not guarantee quality-of-service for real time services.

The data transport is augmented by a control protocol (RTCP) to allow monitoring of the data delivery in a manner scalable to large multicast networks. It partners RTP in the delivery and packaging of multimedia data, but does not transport any media streams itself. The primary function of RTCP is to provide feedback on the quality of service (QoS) in media distribution by periodically sending statistics information to participants in a streaming multimedia session. RTP and RTCP are designed to be independent of the underlying transport and network layers [2].

7. **DESIGN CHARACTERISTICS**

The SIP-based system architecture shown in Figure 1 uses SIP as its call signaling protocol due to the strongpoint described in the scope section above. In this section, design characteristics of the conferencing system are listed below:

1. Besides the scenario that participants can initiatively dial-in to a conference, this conference will also be able to dial-out to users who are already registered in the server. The dial-out list will be determined at the beginning of a meeting or added during the conference by someone (e.g.,
the moderator of a conference) who has right to invite participants.

2. Conference can be reserved at one’s willing time in the future. When the conference begins, it will invite users in the dial-out list.

3. The conference server will distinctively send full or partial conference state information to the authorized participants who have subscribed the notifications with different demands.

4. This multimedia conferencing system will not only support audio or video like traditional products. It will also enable exchange of documents among multiple participants.

8. CENTRALIZED SYSTEM

The proposed conferencing system can be divided into two parts: server and participants. As shown in Figure 2, participants can be personal computers. As in centralized conferencing, the heavy load of media will become the bottleneck of the whole system; the media process function is taken out of original server. Set of media servers are used to share the system load. When a participant enters a conference, a RTP stream is established between the participant and media server.

Participants in a meeting can chat, exchange data using Message Session Relay Protocol (MSRP). MSRP is a protocol for transmitting a series of related instant messages in the context of a communications session. Here, the application instantiates a session with the Session Initiation Protocol (SIP). MSRP is used within a SIP session to do Instant Messaging (IM) or to transfer file attachment [5].

To a certain extent, media server and MSRP server play the same role in the system. Database (DB) server stores the conferences and users information [1].

9. IMPLEMENTATION

JMF (Java Media Framework) will be used owing to the fact that it provides a better and easier way to create applications that present, capture, manipulate, and store time-based media. The framework also enables custom processing of raw media data. The RTP APIs in JMF support the reception and transmission of RTP streams and address the needs of application developers who want to use RTP to implement media streaming and conferencing applications.

**DataSources, Players and Processors are integral parts of JMF’s high-level API for capturing, and processing time-based media.**

After capturing the media packets at the server with the help of RTP, the same will be sent to the client site. As soon as RTP streams arrive at client site, a player will be created so as to play the RTP streams. This player will be a JMF player with proper GUI and control components.

Following steps would be used to implement the system:

1. An interface would be developed for capturing video packets from a webcam and displaying the same on the terminal for a single user.
2. The above implementation would be modified to capture video packets using JMF and transfer the same across multiple users thus developing a conferencing system.
3. After the core implementation of conferencing system is complete, an attempt would be made to add mining structures that will give detailed information about the user, what kind of conferencing he/she does and with which user he/she interacts most of the time.

The output of the 1st step listed above can be seen as follows:

---

Fig. 2: Centralized Conferencing System
10. CONCLUSION
Video conferencing has helped people to communicate more effectively. The early conferencing solutions are based on the H.32x standard. The absence of standards has severely limited widespread deployment due to incompatibilities between different vendor solutions. To resolve this problem, much effort has been made on standardizing the design and implementation of conferencing applications.
Our ongoing multimedia conferencing project is part of this effort, aiming at eliminating incompatibility between different platform-oriented solutions, by exploiting H.32X standards. The implementation of video conferencing part is currently underway.
In this paper, we have presented architecture for a SIP-based conferencing system. It is proposed to implement a variety of new functions to meet advanced requirements, such as inviting "n" number of users to the conference, dial-out method, etc. Finally, it is also proposed to implement mining structures to the conferencing system that will give a brief history of the user.

11. FUTURE SCOPE
On successful completion of the proposed intelligent system, efforts will be made to design the conferencing system for distributed architecture that will support large number of heterogeneous networks and devices.

REFERENCES