

Audio and Video Communication Software Design Based on SIP

Shanshan Peng

School of Information Science and Engineering, Hunan International Economics University,
Changsha, China, postcode: 410205
E-mail: matlab_bysj@126.com

Abstract

This paper makes a design which implements the audio and video communication based on SIP over the Win2K platform. The oSIP and eXosip were used to exchange the Signals, and the jrtp was for packing the audio and video data. In the end, the multithreading was adopted to make the software perform well.

Keywords: session initiation protocol (SIP), oSIP/eXosip, jrtp, audio and video communication, real-time transport protocol (RTP)

Copyright © 2014 Institute of Advanced Engineering and Science. All rights reserved.

1. Introduction

With the rapid development of Internet technology, multimedia communication technology has been rapid development, which audio and video communications are the most interesting development. IP phones (VOIP) is a technology to transmit voice through IP network. In 1995, Vocaltec company launched the first phone system based NTERNET [1]. Since then, IP phones started to get people of all ages. Compared to the traditional PSTN network, VOIP is cheaper and easier to achieve. In recent years, it has developed a lot of IP telephony standard, and its call quality is also rising, it has been already comparable to the traditional PSTN network.

Most of the domestic telephone network is based on H.323 protocol, because the H.323 protocol is proposed by ITU-T, it is based on IP multimedia standards, which are developed by telecommunications network signaling and protocols, rather than it is raised specifically for IP phones, so the IP network and its application are the great complexity, and it is not easily extended. Relative to H.323, SIP protocol has inherent advantages, multimedia communications of networks with SIP protocol is a natural thing. China has developed a true VoIP software based on SIP, which is also rarely, but there are examples of success, such as Tsinghua University NGN laboratory has developed CoolSIP [2]. Voice network communication can not only be achieved by SIP protocol, but also video network communication can be realized, this paper is to introduce SIP-based audio and video softphone principle and implementation.

2. SIP Profile

Session Initiation Protocol (referred to as SIP) is one of the core protocol in next generation network (NGN). It was originally developed by the IETF MMUSIC (Multiparty Multimedia Session control) working group, and a standard was proposed in 1996, the signaling control is solved for IP network communications, SIP is a signaling work at the application layer, it is used to create, modify, and terminate multimedia sessions process [3].

Compared with H.323, SIP has a simple, scalable, and there are the existing Internet application features closely, therefore, in recent years, the development of SIP applications are much faster than H.323. SIP starting point is that IP telephone service network is architected based on the existing Internet. Therefore, SIP has a completely different design ideas with H.323, it is a decentralized protocol, the complexity of the network will be pushed to the edge of the network devices, and it is compared with IP phones based on H.323 recommendation, SIP

needs relative intelligent terminal. For the occasion which the user terminal is non-intelligent terminals, SIP can also be used as a call signaling [4].

SIP protocol [5-9] refers and widely uses the two kinds of network protocol entities: SIP is used for hypertext transfer protocol (HTTP) of web browser and for email Simple Mail Transfer Protocol SMTP. From the beginning HTTP, SIP draws C / S design mode, and the Uniform Resource Locator and uniform resource identifier (URI) are used, SIP draws plain text encoding scheme and header style from SMTP, SIP re-uses SMTP head, such as To, From, Date, Subject, etc.

3. Design Program

SIP call setup functions mainly rely on the completion of various entities. Entity (client) is sent by the SIP to generate a request, it is sent to the receiving SIP entity (server). Server processes the request, and one or more client response message are returned. Corresponding request and response constitutes a transaction (Transaction). In SIP protocol, communication component consists of two parts: the user agents and network servers.

User Agent is an intelligent terminal system, the need joins the call on behalf of clients, which includes two parts: a user agent client UAC (User Agent Client), which is used to initialize a call request; User Agent Server UAS (User Agent Server), which is usually the called destination, it is used to answer calls and to send out a response. Network servers include registration server (Registrar), we can keep abreast of the registered SIP users in the region; proxy server (Proxy Server) is similar to HTTP proxy, which receives a request and send the request; redirect server (Redirect Server) does not submit backwards after receiving a request, but the client area is directly informed to request the next hop server. Figure 1 illustrates the process in the entire audio and video communication mechanism, the Proxy Server is only identified in three network server figure, the customer registration processes are not introduced.

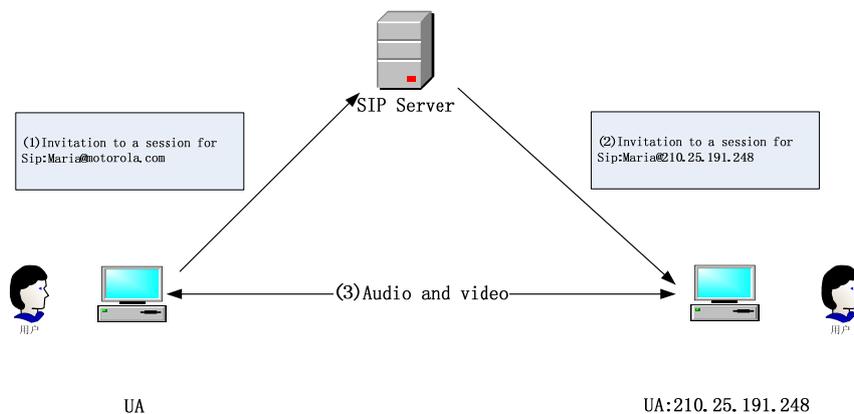


Figure 1. Audio and Video Communication Processes

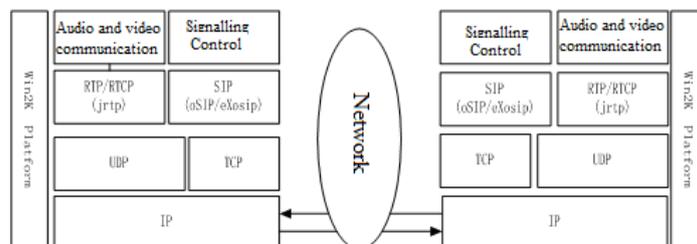


Figure 2. Software Terminal Framework

In the design process, UA is the focus of the design, because it not only responsible for initiating a call, and the call is processed, it is into a SIP entity of Human-Computer Interaction(HCI). The following design is based on a client operating system Win2K (UC) software, the software structure is shown in Figure 2.

SIP softphone structure includes two modules, which are signaling control module and audio and video communication module. Signaling control module is implemented by the SIP protocol, the specific protocol stack are oSIP and eXosip, its main achievement is to create, modify and dismantle the call; audio and video communication module consists of three sub-modules composition, which are audio and video data interface, audio and video codecs and RTP transport, its functions are audio and video capture, encoding, transmission and playback.

3.1. Signaling Control Module Design

Currently, a relatively large number of open source SIP protocol stack includes Vocal, sipX, ReSIProcate and oSIP.

Signaling control section uses oSIP protocol stack, because oSIP source code is mainly one of the few protocol stack which are written by using C language, it has the short and concise characteristics, it focuses on the SIP underlying parsing, and there is more efficient. But there are also disadvantages, the first, the availability is poor, there is no good API package, so that the upper layer application is broken at the time the call protocol stack; secondly, a transaction level protocol parsing process is just done, the resolve of call, session, dialog and other process are lacked, this also increases the application difficulty; again, the mechanisms for concurrent processing thread is lacked, so it has limited processing capability. eXosip is an extension of the agreement set oSIP, which partially encapsulates oSIP protocol stack, it is made easier to use. eXosip increases analytical call, dialog, registration, subscription and other processes, there is more practical. In summary, oSIP protocol stack plus eXosip protocol stack to achieve SIP protocol, this is a good choice.

3.2. Audio and Video Transmission Module Design

Audio and video data interface section includes audio and video capture and playback. Because it is developed based on Win2K system platform, so the Windows API function of oneself library is used in audio capture and playback, waveInXXX class is used with recording function, waveOutXXX class functions is used in play sound. AVICap window class VFW (Video for Windows) is used in Video capture, VFW is developed by Microsoft, which is released with the Windows operating system. A key idea of VFW playback is that no special hardware, which enables the application program digital, and the video is gotten from traditional analog video playback sources.

audio and video encoding Includes audio encoding and video encoding. When audio and video data are captured the amount of data is usually very large, it is not conducive to the transmission network, through coding, without compromising the quality of voice and video, the amount of data are reduced maximumly. Audio codec standards include G.711, G.723.1 and G.729a; video codec standards include H.263 and H.264.

Transmission of audio and video packets use a real-time transport protocol RTP (Real-Time Transport Protocol) and RTCP (Real-Time Transport Control Protocol).

4. Implementation

Based on the above scheme, the client software implementation includes two parts, which are signaling control and audio & video communication, in order to improve the efficiency of resource use, signaling control, and audio & video communications are achieved by using sub-thread.

The relationship between each thread is shown in Figure 3.

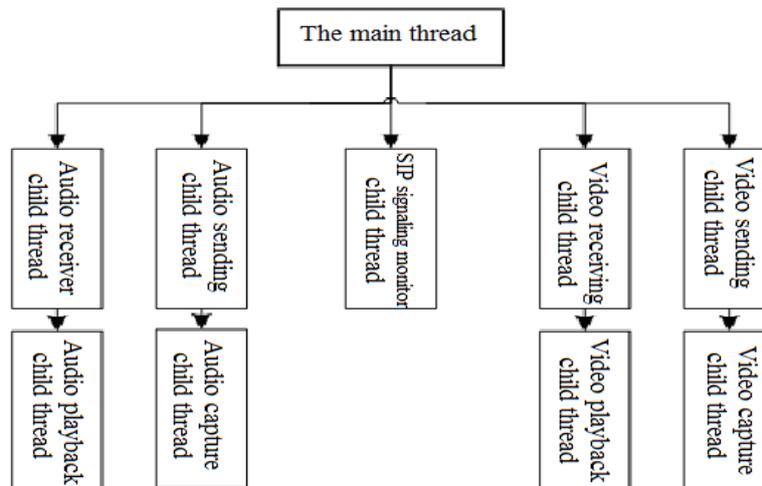


Figure 3. Inter-thread Communication

When the software starts, the main thread is generated, and then nine sub-threads are produced in the main thread. The main thread is used to coordinate the behavior of each sub-thread action, the message notification mechanism is communicated with the child thread, and it also responsible for signaling sending and processing section.

4.1. Signalling Control Module

Signaling control module is responsible for the registration and to initiate a call to the server, mainly in sub-thread monitor of SIP signaling and the main thread processing.

SIP signaling exchange process is the upper application, which calls SIP protocol stack and provides API functions, the protocol stack is notified to appropriate action, the stack will detect the underlying time, which is reported in the form of message to the application layer, after the application layer receives SIP message made, and appropriate treatment is made.

The entire process of registration, call signaling control flowchart is shown in Figure 4.

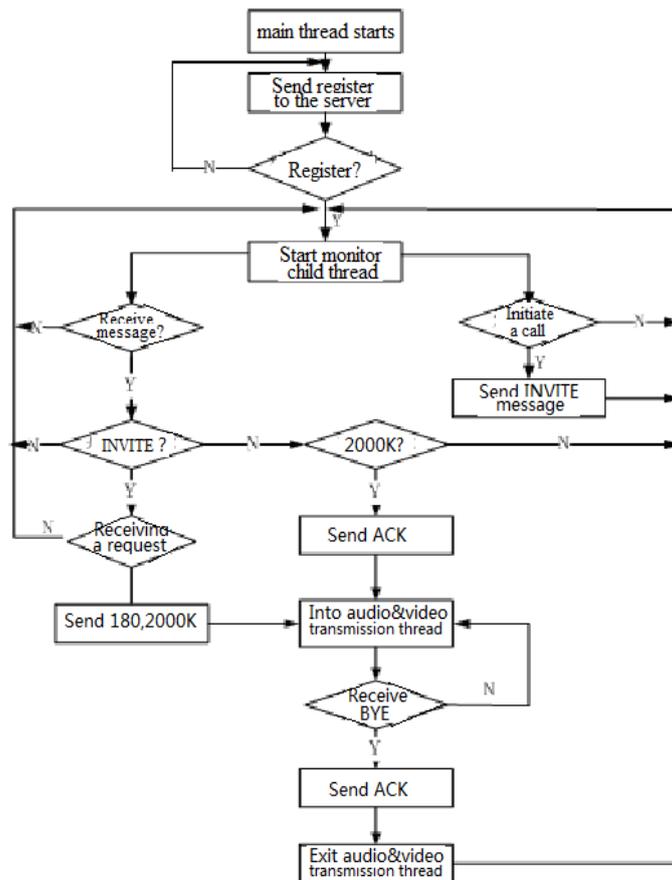


Figure 4. Signaling Call whole Process

4.2. Audio and Video Communication Module

Audio and video data transmission thread of communication module includes video and audio data sampling, coding and RTP packaged and sent transmission; receiving thread includes RTP packets receiving, decoding and audio & video playback. Encoding and decoding protocol standards are used in audio and video communication process, during the the SIP signaling exchange process, agreement is reached through negotiation of message body.

Audio and video communication specific data processes are shown in Figure 5.

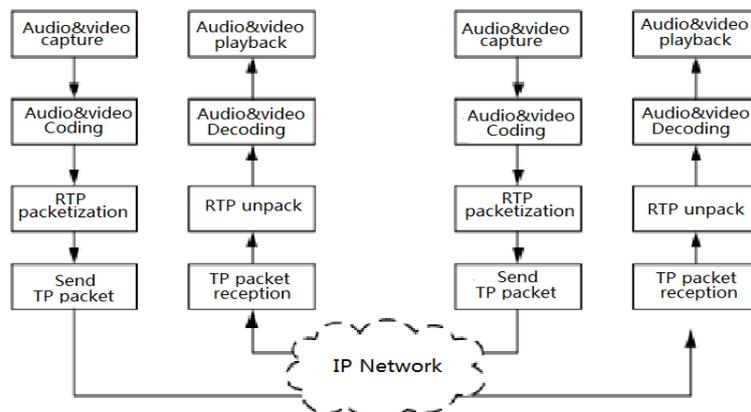


Figure 5. Audio and Video Communication Process

Among them, the audio codec standards include G.711, G.723.1 and G.729a; video codec standards include H.263 and H.264. What kind of criteria is used in the specific call? the call signaling exchange is negotiated between the two sides.

5. Conclusion

Audio and video phone soft design is proposed based on SIP Protocol, the design ideas and methods are detailed to change the software, and these are implemented by using Visual Studio on Win2K platform. In future studies, we will consider the security mechanisms[10], and the P2P technology is used, the part function of the server will be pushed to the brink of the network, the pressure is reduced on the server. SIP protocols are for its simplicity, versatility, portability, and other characteristics, which are attention by everyone[11-13]. Development and Utilization of SIP multimedia communications will be the future development trend of network communications.

References

- [1] BUR GOODE. *Voice Over Internet Protocol*. Proceedings of the IEEE. 2002; 90(9): 1495-1517.
- [2] Cao Xuanming, Zhang Sifa, Huang Yongfeng. Base on SIP Instant Message Integrate Application. *Data Communications*. 2007; 6(3): 33-37.
- [3] SI DuanFeng, HAN XinHui, LONG Qin. A Survey on the Core Technique and Research Development in SIP Standard. *Journal Of Software*. 2005; 16(2): 239-250.
- [4] LI Jun, XIE Zanfu, CUI Huailin. Design and Implementation of Voice Communication Based on SIP Protocols. *Computer Engineering*. 2005; 31(24): 117-119.
- [5] Bai Jianjun, Peng Hui, Tian Min. *SIP Revealed*. People's Posts and Telecommunications Press, Beijing, China. 2003; 6.
- [6] Rosenberg J, Schulzrinne H, Camanilo G. SIP: Session initiation protocol. RFC 3261[EB/OL].<http://www.ietf.org/rfc/rfc3261.txt>,2011.
- [7] Elwell J. Connected identity in the session initiation protocol (SIP),RFC 4916[EB/OL].<http://www.ietf.org/rfc/rfc4916.txt>,2011.
- [8] Johnston AB. *SIP: Understanding the Session Initiation Protocol*[M]. Boston,USA:Artech House,2004.
- [9] Handley M, Jacobson V, Perkins C. *SDP: Session Description Protocol*. 2006.
- [10] Arkko J, Torvinen V, Camarillo G. Security mechanism agreement for the session initiation protocol (SIP), RFC 3329[EB/OL].<http://www.ietf.org/rfc/rfc3329>,2011.
- [11] Liu Shenxiao, Wang Xuechun, Wang Zhigang. The Achievement of VoIP's SIP Calling in Tri-Play Switch. *Bulletin of Science and Technology*. 2013; (10)
- [12] XU Pengyu, XU Zican. Design and Implementation of Monitoring System Based on SIP. *Computer Engineering*. 2013; (11).
- [13] Peterson J, Jennings C. Enhancements for authenticated identity management in the session initiation protocol (SIP),RFC 4474[EB/OL]. <http://www.ietf.org/rfc/rfc4474>,2011.