Design and implementation of VVOIP telephone based on Android

Xiangdong You[1, a *]

[1]Beijing University of Posts & Telecommunications, China

[a]youxiangdong@bupt.edu.cn

Keywords: Android VVOIP ARM SIP   H264

Abstract. This paper proposes a design of VVOIP (Video and Voice over IP) on smart devices, with ARM Cortex A8 CPU structure in hardware, Android OS and JAVA language in software and C in Native layer. In the VVOIP, SIP is adopted as session protocol, and RTP packages carry H264 video streamer and GSM audio streamer.

Introduction

With rapid development of mobile Internet in recent years, smart devices based on mobile terminal change with each passing day. Among operating systems of intelligent terminals, IOS and Android OS account for a large proportion in the consumer electronics market. As traditional voice-only calls have been unable to meet consumers' needs, smart terminal based videophone has been accepted by more and more people, especially by households and enterprises.

It is not hard to understand that video calls should meet higher requirements than voice calls, such as more powerful CPU processing capacity and higher bandwidth. However, with the booming of semiconductor chip industry and the popularity of broadband Internet, both these two key issues can be overcome. The design of the VVOIP phone based on Android proposed in this paper can provide smoother video calls.

The principles of VVOIP

VVOIP is a high-performance network terminal on overall, using ARM Cortex A8 Processor. Soft-switch technology is adopted for the call part. Call process is achieved mainly by SIP and RTP protocols, which are responsible for the set-up of the call and the interaction of data during calls separately. Another feature of video phone is the acquisition, compression and transmission of multimedia data. We used H264 hard coding in video encoding part, while VP8, H264[1], MPEG4 soft coding is also supported.

SIP   Session Initiation Protocol (SIP) was developed by the IETF in 1999. It is mainly used for the setting up、 modifying and releasing of the session in network communication. In another word, SIP manages session by solving most of the problems in the setting-up of the session, simplifying the application layer software.[2] Just like HTTP, SIP is also a plain text protocol, which makes it possible to work among different systems. SIP belongs to the session layer in OSI, over the transport layer. SIP is widely used in multimedia communication, and it is much more popular now with the development of multimedia communication technologies.

Features of SIP protocol[2] include convenience, easy development, flexibility, suitability for various applications and compatibility with other technologies.

RTP   People are sensitive to audio data in the multimedia communication. Both the loss of a data frame and a long delay will result in noise and will lead to decline in user experience; while for video data, the missing of few frames or a delay will only cause the image mosaic, which does not have much influence in user experience. From this we can see that audio and video has different requirements for network transmission in multimedia communications. There are two transmission modes in IP network, one is the connection-oriented TCP, and the other is the connectionless UDP. UDP cannot

guarantee the reliability of the audio data, while TCP's three-way handshake will cause unnecessary delay. What's more, TCP can't realize retransmission.

Through this analysis we can find that the original transport-layer protocol cannot support multimedia data transmission. Out of this demand, the International Organization for Standardization has developed a set of RTP/RTCP Protocol for transmission of multimedia data streams. .

Video coding technique   H264 is a new video compression standard developed by a joint video team which consists of ISO/IE and ITU-T. H264 has three subsets: basic subsets, subject subsets and extensions. Basic subset was designed specifically for video conferencing applications, which provides powerful error cancellation technique and supports low-latency encoding/decoding technology making the display of video conferences much more natural.

   H264 puts forward two new concepts: Video Coding Layer (VLC) and Network Adapter Layer (NAL). The former is a description of the compression of video content information, which defines the syntax of the macro block layer and layers below; while the latter is a description of transmission through specific network, which defines the syntax of the slice layer and layers above. This structure makes information encapsulation much easier and provides better priority control of information to adapt to different network applications. By adopting H264 video coding standard for network transmission, high compression ratio and stronger robustness can be achieved. According to RFC3984, there are three types of subcontracting modes when using RTP to carry H264's NAL unit: Single NAL unit mode, Interleaved mode and Non-interleaved mode. [4]

   All forms of the packets of data extraction units in H264 are called simple package. This packet strategy is to head a NAL packet into a RTP packet before sending it. H264 video streams should be packaged into RTP packets following the RTP specification when transporting over RTP. According to the encapsulation of RTP specification [5], and to avoid data segmentation in IP layer, NAL units generated in video encoding procession should not be larger than the maximum size of transmission unit.
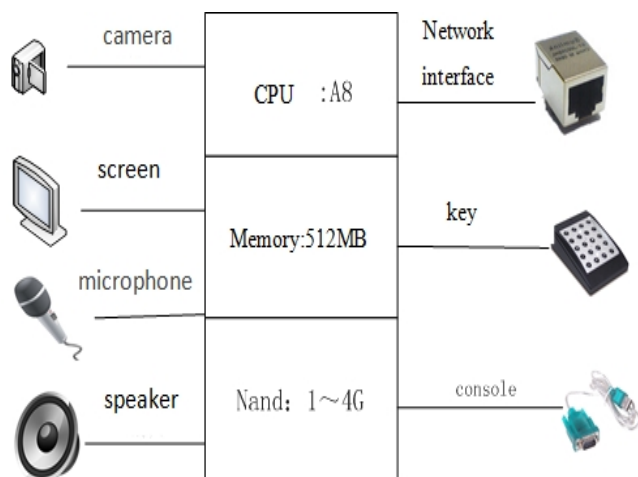
Hardware design of VVOIP

Fig.1 Hardware frame of VVOIP   Fig.2 Prototype of VVOIP

Hardware diagram of VVOIP   Its core circuit adopts S5PV210 produced by Samsung as core processors. This processor was developed based on ARM Cortex-A8 processors, and its clock can speed up to 1GHz. S5PV210 contains an integrated graphics engine called PowerVR SGX540. This VVOIP was configured by a memory of 512MB and a FLASH memory of 2GB, and its function module contains a network interface, a keyboard, CONSOLE interfaces, digital cameras, LCD monitors, microphones, speakers and so on, as shown in Figure 1.

Interfaces of VVOIP   Different from the traditional telephone, smart video phone based on Android has a more complex structure, so as to cover more function modules and provide various interfaces to meet more demands. In addition to the basic handle interfaces, it should also provide interfaces to

Ethernet, USB, LCD and so on. Based on the functional and performance requirements of VVOIP, we designed more than one telephone interfaces as shown in table 1.

Interface	Performance	Number	Description
Wired networkRJ45	10/100Mbps	1	One RJ45 socket
Camera	CMOS camera	1	VGA or above
Monitor	Seven inches TFT LCD	1	800x480 or above
Key	Basic functional keys for Android	5	One switch/standby key and five common key named K1-K5
USB HOST	Host USB 2.0	1	One USB host socket
USB OTG	OTG USB 2.0	1	Used for ADB for Android
LEDs	Red LED for power Green LED for network	1-4	LED for network is put on RJ45 socket
SD card socket	SD card	1	Mini SD-card socket
Phone handle	Including one speaker and one MIC	1	four core handle socket
Hands-free telephone circuit	One Speaker of 8Ω and 1W ,one MIC	1	Speaker(8Ω 1W)
power	5V/2A	1	220V/5V power adapter
Insert spring of handle	Pick-up machine testing	1	Provide pick-up machine detection ting handles

Table 1: Interfaces of VVOIP

Software design of VVOIP

Linphone Linphone is an open source VoIP software supported by many platforms such as Android 、ISO 、Linux and so on. Linphone uses hierarchical design so that the programmers can transform it as they want based on their system characteristics to develop customized products. Linphone uses SIP protocol as the call signaling, and transports media streaming via RTP/RTCP, which support the communication among different software using the same standard protocol. As for the design proposed in this passage, it is a must to reform Linphone in order to improve the efficiency of the telephone. The reformation, such as replacement of hardware and software encoding, will promote the operating efficiency of the entire system.
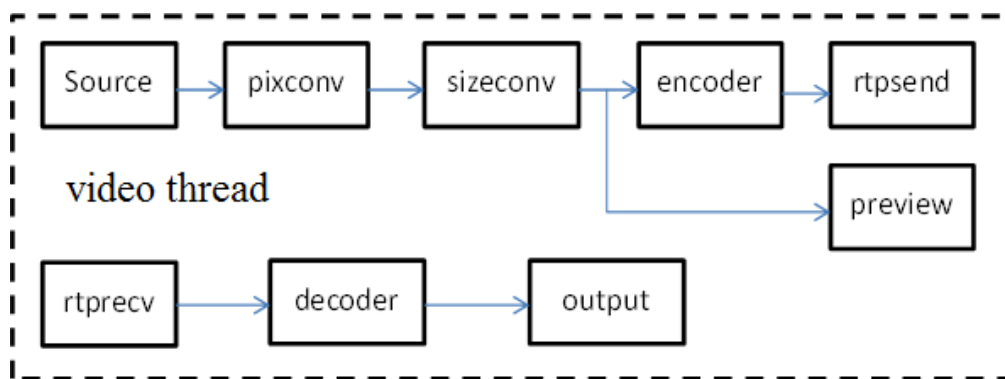

Figure 3: Mediastreamer2 video thread

Linphone adopts Mediastreamer2 as the multimedia processing module, which is the core module of Linphone's streaming media handling module. The capturing of audio and video massages , the sending and receiving of RTP streaming of audio and video massages, the encoding and decoding of audio and video massages and the playing of audio and video massages are all accomplished in this module. Its core work is completed by three threads: video stream processing thread (Figure 3), audio processing thread as well as video capturing thread. Filter structure is the core interface of Mediatreamer2. A

Filter object is a process, and a complete streaming media process will be built up after connecting the different Filters. Filters transport data directly through a data queue.

Transplantation and transformation of Linphone software   Compilation of Linphone library. In order to achieve better performance of H264 hardware encoding, you need to call a third party library file and to replace the original API of H264 by a new API which supports hardware encoding. On Android platform, H264 of Linphone is a separate module. In order to obtain better results, an overall optimization should be given to this part. First of all, d a referenced third-party library should be added in the Android.mk file. Compile the file by using NDK(Native Debug Kit) on Android platform, after getting the new H264 library file that supports hardware encoding. Don't forget to place the new library file to the place specified by the system to make the new library file accessible for applications of upper layer, and then hardware video encoding can be realized.

Prototype of VVOIP

As is shown in figure 2, this VVOIP adopts H.264 for video processing and SPEEX coding standard for audio processing during two-way communication. This phone works well, and indicators of the prototype are listed as follows:

      Video encoding/decoding: VP8, H.264 and MPEG4 supported

      Video resolution: QCIF, CIF and VGA

      Frame rates: no less than 15fps, adaptable according to bandwidth

      Video bit-rate: 280-330kbps (CIF, 15fps)

      Audio encoding/decoding: G.711u, G.711a, G929, GSM and SPEEX supported

      Network interfaces: 10M/100Mbps cable RJ45, 802.11g Wireless and WiFi

Conclusion

This paper proposed the design and implementation of a visual VOIP based on Android and ARM Cortex-A8, which uses SIP protocol as call signaling. This telephone supports registration, call and video call, and adopted RTP Protocol to transmit call media streaming. Among this, its audio encoding supports G711A, G711U, G722, GSM, SPEEX 8K and SPEEX 16K, and its video encoding supports H264, VP8 and MPEG4.

    The design ideas of the VVOIP proposed in this paper can be used in in-vehicle wireless video monitoring, network monitoring and many other fields. In a word, it can be widely promoted, and has tremendous commercial value in the internet of things in the future.

References

[1] Yousef O. Sharrab and Nabil J. Sarhan, Detailed Comparative Analysis of VP8 and H.264, 2012 IEEE Symposium on Multimedia, pp. 133 - 140, 2012.

[2] Zhang Youbo, Zhang Huanqiang and Sun Liming, Design and implementation of video conference system based on SIP, Computer Engineering,Vol 31, Nov 2005 In Chinese

[3] Zhang Haijun, Zhang Jianjun , Yang Yingeng and Wu Kejie, Design and implementation of RTP transmission control and real-time video monitoring system, Microcomputer and Application vol 28

（5）, 2009. In Chinese

[4] Li Xin, Network video transmission based on H.264, Beijing University of Posts and Telecommunications,2006 In Chinese

[5] Schulzrinne H. Casner S. RTP: A Transport Protocol for Real-Time Application, RFC3550, 2003.