Concept of spoken dialog system based on voice over IP telephony

Matúš Pleva, Jozef Juhár, Anton Čižmár

Department of Electronics and Multimedia Communications,
Technical University of Košice, Park Komenského 13, 041 20 Košice,
Slovak Republic
Email: Matus.Pleva@tuke.sk, Jozef.Juhar@tuke.sk, Anton.Cizmar@tuke.sk

Abstract

The problems of internet telephony (VoIP - Voice over IP) applications are widely discussed nowadays. New protocol standards allow us to use the internet as the main communication media for spoken dialog system access. This article describes a solution for spoken dialog system based server with PSTN connectivity using only software based VoIP clients integrated. This solution makes the spoken dialog systems more attractive for developers, because it needs only a PC with internet connectivity and the dialog services could be reachable from the PSTN network without expensive telecommunication hardware.

1. Introduction

The building of spoken dialog system server is not so easy nowadays because of the expenses for telecommunication hardware which is more expensive then the PC hardware at all. The PSTN or digital PCI cards, GSM gateways, VoIP gateways makes the system accessible for all regional telephone users, but there are always costs for PSTN line or GSM card also. Compared to the VoIP access connectivity, which is mainly for free or for very low monthly costs. The VoIP spoken dialog server is also movable and could be placed in different location with internet connectivity when needed. The problem is to choose one of the VoIP standards, for building the fully functional AudioServer (part of spoken dialog system) with VoIP client integrated. The integrated VoIP client could provide: call control, caller identification, call-back or instant messaging (information about the dialog result – weather, train connection, etc.). This client opens also a possibility of multi-user handling of the spoken dialog server.

2. Choosing a VoIP standard

First of all we need to choose which VoIP standard is the best for integrating to the spoken dialog system AudioServer. The goal is to choose the most frequently used standard between VoIP providers, with open source or non-proprietary protocols and clients and with a possibility to build a multi-user server.

2.1. SIP

The Session Initiation Protocol (SIP) is a signalling protocol for call setup and teardown including video, voice and instant messaging [1]. This protocol is the most frequently used standard between cheap VoIP providers.

There are also a lot of complete open source clients for many platforms. The SIP protocol is transport-independent, because SIP can be used with UDP, TCP, SCTP, etc. The protocol is text-based (built on principles of HTTP and SMTP), allowing for humans to read and analyze SIP messages.

The problem is NAT (network address translation) traversal and firewalls. SIP uses a new UDP connection for every call. There are o lots of possibilities of solving the problem: Universal Plug and Play (UPnP), STUN (Simple Traversal of UDP through NATs), Connection Oriented

Media (Comedia), RTP Relay (TURN), Application Layer Gateway (ALG) or different tunnelling protocols. Anyway this is a weakness of SIP protocol.

2.2. H323

The H.323 Call Signaling protocol is based on the Q.931 protocol and is suited for transmitting calls across networks using a mixture of IP, PSTN, ISDN, and QSIG over ISDN. A call model, similar to the ISDN call model, eases the introduction of IP telephony into existing networks of ISDN-based PBX systems, including transitions to IP-based Private Branch eXchanges (PBXs) [2].

This protocol is good for implementing VoIP feature to existing PSTN network, but VoIP providers do not use it because of binary based protocol with much more complicated implementation in open source environment.

2.3. Skype

Proprietary VoIP protocol of Skype network could be used because of the possibility of building an AudioServer plug-in to existing Skype clients [3]. But there could be only one client running on one PC server, and the clients are proprietary only.

In this case also costs for incoming (dial in) number will increase the monthly expenses. But it is a good way to make the dialog system accessible for internet users, because Skype is frequently used free VoIP network nowadays.

2.4. IAX2

IAX is the Inter-Asterisk eXchange protocol native to Asterisk PBX (free open source PBX software for PC). It is used to enable VoIP connections between servers as well as client-server communication.

IAX2 uses a single UDP data stream to communicate between endpoints, both for signaling and data [4]. This is a big advantage for communicating trough firewalls, NAT servers etc.

Disadvantage is that not all open source SIP clients and VoIP providers enable this protocol. Because a lot of VoIP providers use SIP protocol as the main option and IAX as a possibility.

3. Building the solution

The developed spoken dialog server in Slovak language is built on Galaxy architecture using Intel Dialogic telephony card as the telecommunication interface (see Fig. 1) [5]. On the Dialogic interface [6] there is one PSTN line connected, 3 GSM and one Skype gateway.

Then a VoIP (SIP) gateway (with 2 FSX ports) was also purchased for testing SIP clients from various platforms and the open source client chosen for integration to Galaxy-based AudioServer [7].

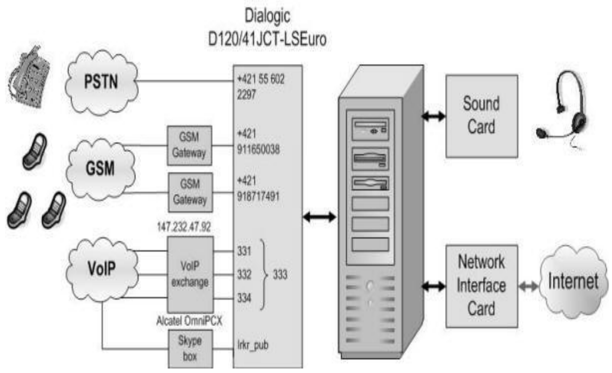

Figure 1. Spoken dialog server based on Galaxy architecture using Dialogic telephony card as a telecommunication interface

The chosen open source VoIP architecture was pj-sip [8]. It’s a complete SIP protocol stack with high performance and extremely portable source code. This small VoIP client was tested on Windows, Windows Mobile (for calls from PDA devices) and Linux platform.

The next step is the VoIP client integration to Galaxy based AudioServer, which is responsible for audio data exchange between the spoken dialog server (see Fig. 2) and the SIP provider.

The audio data are transferred using a small memory buffers and the codec used during the SIP transmission could be chosen from available codec list from Table 1, but both clients needs to agree on chosen codec before the audio transmission starts. Every SIP client needs a registration to some VoIP provider.

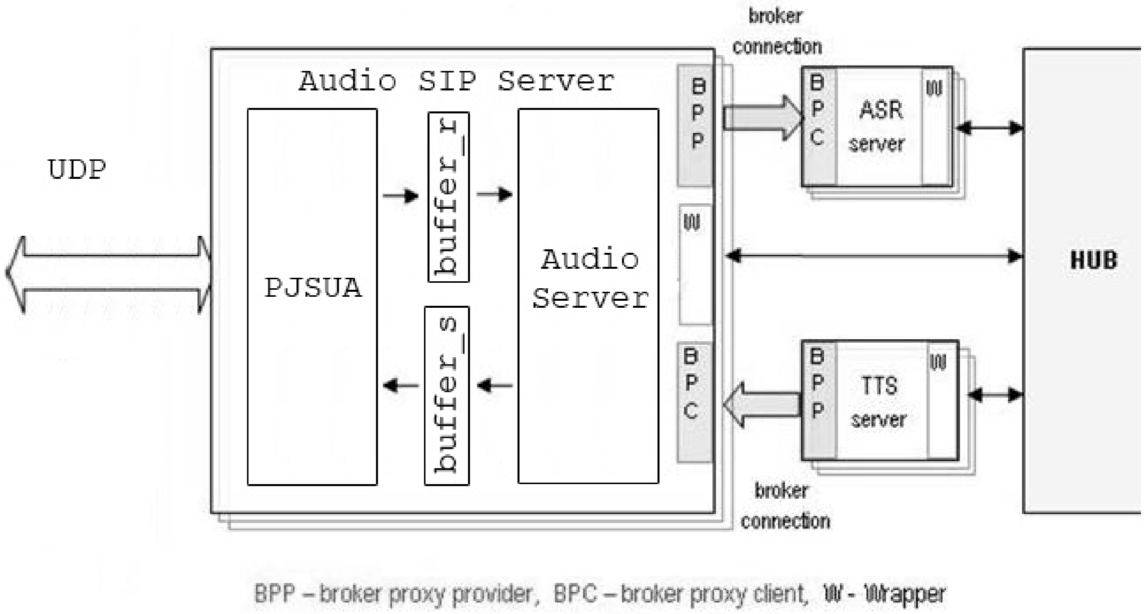

Figure 2. AudioServer based on Galaxy architecture using pjsua SIP client as a telecom. interface

"Table 1. Available codecs in chosen pjsua SIP client."

Codec	Bitrate [kbps]
G711	64
G722	48-64
GSM 06.10	13
iLBC	13.33-15.2
L16	128
Speex	2-44

4. Conclusion and future plans

The AudioServer built on SIP client opens up the possibility to build and test the spoken dialog server using only a computer with internet connection. No additional hardware is necessary for building a multi-user spoken dialog server interface on one PC.

Thanks to a low cost Slovak VoIP provider with a Slovak dial in number the server could be available with general access from all telecommunication networks.

In the future we plan to build an AudioServer with built-in Skype plug-in, which will help to connect the spoken dialog server applications to a free and most popular VoIP network in Slovakia. The current server is connected to the Skype using additional Skype gateway and Dialogic card (see Fig.1).

8. Appendix and acknowledgments

The work presented in this paper was supported by the Ministry of education of Slovak Republic under research projects AV 4/0006/07, MVTS COST2102/07, VEGA 1/4054/07 and AV 4/2016/08 and Slovak Research and Development Agency under research project APVV-0369-07.

References

[1] http://en.wikipedia.org/wiki/Session_Initiation_Protocol

[2] http://en.wikipedia.org/wiki/H323

[3] http://www.skype.com/intl/en/

[4] http://en.wikipedia.org/wiki/IAX2

[5] Juhár, S. Ondáš, A. Čižmár, M. Rusko, G.Rozinaj, R. Jarina, "Development of Slovak GALAXY/VoiceXML Based Spoken Language Dialog System to Retrieve Information from the Internet", *Proc. of ICSLP*, Pittsburgh, USA, September 17-21, 2006, pp. 485-488

[6] http://www.dialogic.com/default.htm

[7] http://communicator.sourceforge.net/

[8] http://www.pjsip.org/