International Conference on Communication Theory, Reliability, and Quality of Service

Past, Present and Future of IP Telephony

Vasco N. G. J. Soares[1,3], Paulo A. C. S. Neves[1,3], and Joel J. P. C. Rodrigues[2,3]

[1] Superior School of Technology, Polytechnic Institute of Castelo Branco, C. Branco, Portugal

[2]Department of Informatics, University of Beira Interior, Covilhã, Portugal

[3]Institute of Telecommunications, Networks and Multimedia Group, Portugal

vasco_g_soares@est.ipcb.pt; pneves@est.ipcb.pt; joel@ubi.pt

Abstract

Since the late 90's IP telephony, commonly referred to as Voice over IP (VoIP), has been presented as a revolution on communications enabling the possibility to converge historically separated voice and data networks, reducing costs, and integrating voice, data and video on applications. This paper presents a study over the standard VoIP protocols H.323, Session Initiation Protocol (SIP), Media Gateway Control Protocol (MGCP), and H.248/Megaco. Given the fact that H.323 and SIP are more widespread than the others, we focus our study on them. For each of these protocols we describe and discuss its main capabilities, architecture, stack protocol, and characteristics. We also briefly point their technical limitations. Furthermore, we present the Advanced Multimedia System (AMS) project, a new system that aims to operate on Next Generation Networks (NGN) taking the advantage of its features, and it is viewed as the successor to H.323 and SIP.

1. Introduction

Since the late 90's that terms like Voice over Internet Protocol (VoIP), Internet Telephony and IP Telephony invades the way we communicate over the Internet. These terms may seem equal at first glance, but they encompass some differences.

The first term, VoIP, translates into passing phone calls over a packet data network, without bounds to network type or topology. As a result, VoIP allows phone calls from over Local Area Networks (LAN) up to the global Internet. The term Internet Telephony comes from passing calls across the Internet, and it may use VoIP, but also specific proprietary hardware or even computer software. Finally, the term IP Telephony denotes converged data and voice, and uses VoIP technology. Such integration can lead, for instance, to seamless use of both voice and data messaging.

VoIP is basically the means to grab audio and video in digital form, divide it in small chunks that can be transferred through the network as packets. After, it reassembles the chunks on the other side in a convenient way so that people having a phone conversation have the idea of a circuit switching ordinary call, as shown in Figure 1.

VoIP presents a shift on corporate voice communications where the traditional Private Branch eXchange (PBX) based systems were used to provide internal cost-free communications and sharing of external telephone lines. With VoIP, the PBX gives place to a gateway router and a server on the computer network that controls all calls.

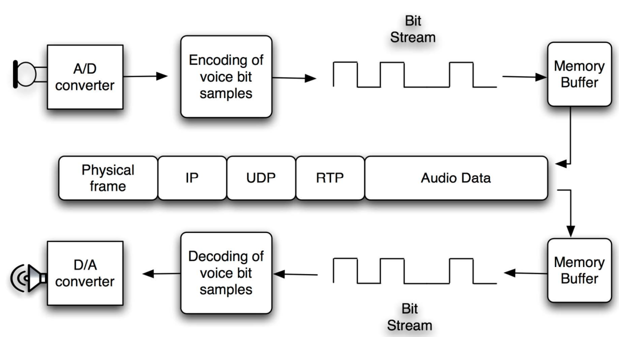

Figure 1. Basic VoIP functional architecture

If we analyze the beginning of the telephony service back in 1876 when Alexander Graham Bell achieved the first phone call with his assistant, up to now there is not much difference for the end user in terms of voice communication. The main advances are in the way such information is transmitted, the quality of speech, introduction of wireless communication and smaller, faster and service-rich terminals.

The first attempt to transmit voice over a packet network came with the Network Voice Protocol (NVP) [1] in the ARPANET. In 1995 a company named

978-0-7695-3190-8/08 $25.00 © 2008 IEEE
DOI 10.1109/CTRQ.2008.12

Volcatec (still working today on VoIP technologies) introduced the first Internet phone software called "Internet Phone". The phone used a home computer (an Intel® 486 processor at 33MHz) with a sound card, speakers, microphone and a modem. Among the limitations were the need for both users to use the same software and hardware, and sound quality was nowhere near the conventional equipment at that time.

In 1996 United States (US) telecommunication companies ask the US congress to ban Internet Phone technology. By 1998 VoIP has reached some potential, with some gateways that allowed computer-to-phone and phone-to-phone IP solutions. Nowadays, for households, broadband Internet access is available in the majority of developed and in-development countries, empowering the use of VoIP significantly. Also, instant messaging systems that combine voice, video and messaging are very popular. The most important operating systems provide support for VoIP communications and Microsoft®, Apple® and Linux® distributions all provide client software and realize the potential of instant messaging communication.

IP telephony brings many advantages over the traditional Public Switched Telephone Network (PSTN) system, namely in terms of cost. Internet access is not billed according to user location, as a PSTN system. So, although the billing parameter is still call time for connections between VoIP and PSTN systems, distance is no longer a primary billing factor.

Another great advantage is flexibility. For PSTN to roll out a new service, it needs to reprogram the entire network of smart circuit-switched systems, while a new service in IP telephony may be as easy to set as the development of a new software feature.

One of the most important parts of a telephone call is the establishment of the call itself. In a packet switched network this is accomplished by a protocol that performs signaling. This paper addresses such protocols, namely the well spread H.323 and Session Initiation Protocol (SIP). However, these protocols are over a decade old and a new approach is also referred here: the Advanced Multimedia System (AMS) project.

The remainder of this paper is organized as follows. Section II shows an overview over the standard VoIP protocols. Section III presents a comparative study of the principal protocols and Section IV provides the conclusions.

2. Voice-over-IP Protocols

IP telephony communication services depend on the use of signaling protocol stacks to set up and tear down calls, to negotiate capabilities, and to carry information required to locate users.

This paper discusses the following standard protocols: H.323, Session Initiation Protocol (SIP), Media Gateway Control Protocol (MGCP), and H.248/Megaco. A brief overview on proprietary protocols, such as Skype, JAJAH, Gizmo, Skinny Client Control Protocol (SCCP), MiNET, CorNet IP and Inter-Asterisk eXchange (IAX), is also considered. Next subsections will present main characteristics, architecture and protocol stack of each one.

2.1. H.323

H.323 is the first international multimedia communications protocol standard. It was published by the ITU Telecommunication Standardization Sector (ITU-T) in February 1996 and its current version H.323v6 was approved on June 2006 [2].

H.323 allows the convergence of voice, video, and data on packet networks. It features World Wide Web and Internet integration, together with PSTN interfacing. Furthermore, it provides diverse applications such as wholesale transit of voice, pre-paid calling card services, residential/enterprise voice and video services. Remote users can perform a video call and simultaneously edit a document in real time over the Internet. H.323 goes beyond, allowing phone or phone services customization, user location, call transfer, or other tasks taking advantage of the HTTP interface between the client/server on the network.

The H.323 architecture is based on the following elements shown on Figure 2: Terminals, Multipoint Control Units (MCUs), Gateways, Gatekeepers, Peer and Border Elements [3, 4]. The Terminals represent the end device of the connection and can be telephones, video phones, IVR devices, voicemail systems, or soft phones. The Multipoint Control Units are used for multiparty conferencing. The Gateways interface the H.323 network to other voice and video networks such as PSTN or H.320. The Gatekeepers are an optional component that is essentially used for call admission and address resolution. A Gatekeeper can allow a call to be placed directly between endpoints (Terminals, MCUs or Gateways) or it may route the call signaling through itself. Peer Elements exchange addressing information and participate in call authorization inside administrative domains and between them. They can be co-located with a Gatekeeper, and may aggregate address information reducing the volume of routing information. Finally, the Border Elements exist between two administrative domains and can assist in call authentication or authorization. They are a particular type of Peer Element.

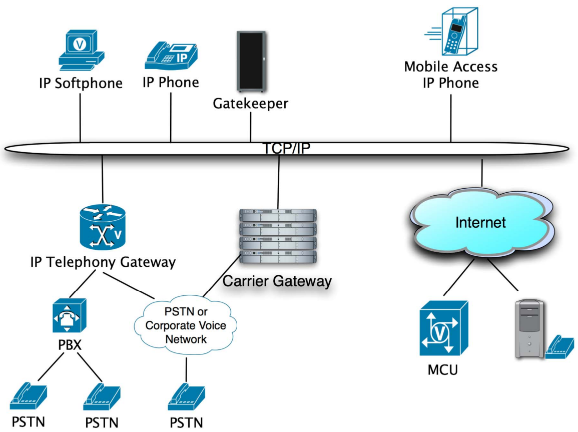

Figure 2. H.323 architecture

According to [5], H.323 protocol stack presented on Figure 3, is composed of many different protocols (ITU-T standards). H.225.0 defines call signaling. H.225.0/RAS is used for registration, admission and status. H.245 is the control protocol for multimedia communication. T.120 is a protocol suite for data conferencing. G.7xx is a series of audio processing protocols. H.26x is a series of video processing protocols. H.235 provides security. H.450.x is a series of supplementary service protocols. H.460.x is a series of version-independent extensions to the base H.323 protocol. Real-Time Transport Protocol (RTP) and RTP Control Protocol (RTCP) are used for media transportation.

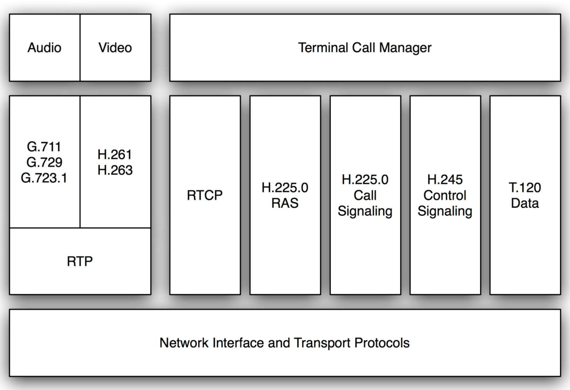

Figure 3. H.323 protocol stack

2.2. SIP

The Session Initiation Protocol (SIP) is an Internet Engineering Task Force (IETF) standard designed for initiating, maintaining and terminating interactive communication sessions between users. These sessions may include voice, video, instant messaging, interactive games, and virtual reality.

SIP's first draft was published in 1996 and the first recognized standard was published later in 1999. The most recent specification was published in RFC 3261 [6] on June 2002.

SIP functions allow user location, user availability, endpoint capabilities, and session setup and session management [3]. It enables voicemail and unified messaging, context-aware communications and location services, integration of communications and applications, and Internet conferencing and collaboration.

The SIP architecture is based on the following elements shown on Figure 4: User Agent (UA), User Agent Client (UAC), User Agent Server (UAS), Proxy Server, Redirect Server and Location Server [4]. The UA is a SIP network terminal (SIP telephone or gateway to other networks) and contains the UAC and the UAS. The UAC is the client in the terminal that initiates SIP signaling, while the UAS is the server in the terminal that responds to the SIP signaling from the UAC. The Proxy Server receives connection requests from the UA and transfers them to another proxy server if the particular station is not in its administration. The Redirect Server receives SIP connection requests and sends them back to the requester including destination data instead of sending them to the calling party. Finally, the Location Server receives registration requests from the UA updating the terminal database with the information about UAC location. It also provides this information to Proxy and Redirect servers.

SIP should be used in conjunction with other protocols to build the complete multimedia architecture presented on Figure 5. The Real-Time Transport Protocol (RTP) is used for transporting real-time data. The Session Description Protocol (SDP) is used to describe multimedia session. The Session Announcement Protocol (SAP) is used to advertise multicast sessions. The basic functionality and operation of SIP does not depend on any of these protocols.

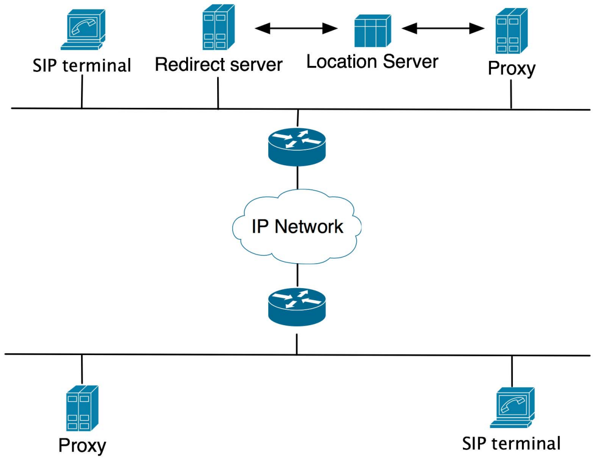

Figure 4. SIP architecture

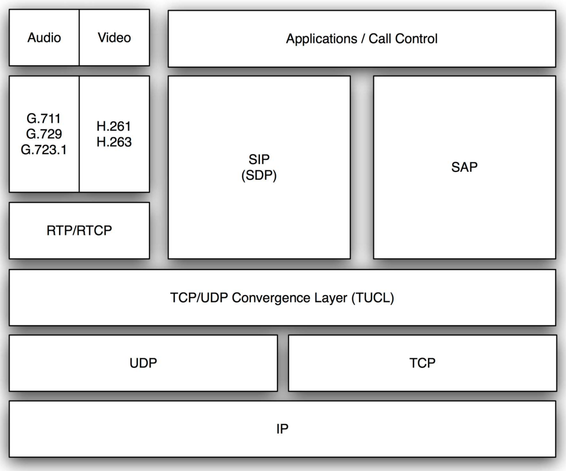

Figure 5. SIP protocol stack

2.3. MGCP

The Media Gateway Control Protocol (MGCP) is an IETF VoIP protocol destined for residential gateways, IP phones and large-scale trunk gateways. MGCP latest specification was published in RFC 3661 [7] on December 2003.

MGCP is used between elements of a decomposed Multimedia Gateway, which consists of a Media Gateway Controller (the Call Agent) that contains the call control "intelligence", and a Media Gateway that contains the media functions. It supports a centralized call control model.

MGCP is not a peer-to-peer protocol, since Call Agents are located at the edge of a network and communicate using a call control protocol such as H.323 or SIP [3]. As shown in Figure 6, it is an edge protocol, so SIP or H.323 still needs to be employed in the core network.

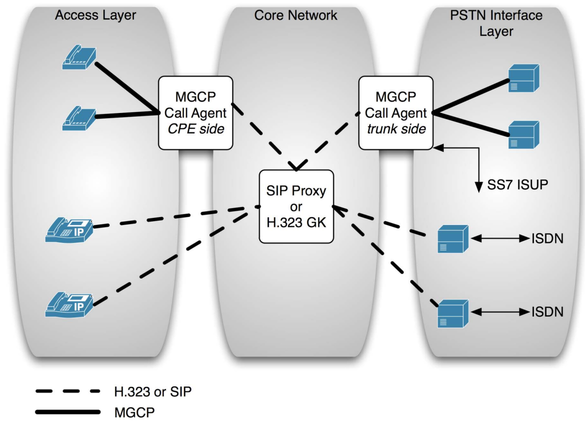

Figure 6. MGCP architecture

2.4. H.248/Megaco

H.248 (or Megaco) protocol represents a joint cooperative effort of the IETF and the ITU-T SG16.

It is the standard for allowing a Media Gateway Controller to control Media Gateways, and it is similar to MGCP from an architectural standpoint and the controller-to-gateway relationship.

This protocol is considered complementary to H.323 and SIP, because a Media Gateway Controller will control Media Gateways using H.248, but will communicate between one another thru H.323 or SIP [4].

2.5. Other protocols

The main goal of this study is to provide a high-level overview of the evolution and architecture of the standard VoIP protocols. Furthermore, we extend this work making a short presentation of other alternative proprietary protocols.

Skype [8] is both a product and the name of a company. As a product, Skype uses a proprietary VoIP protocol and overlay network. It operates on a peer-to-peer model, rather than on the more traditional client-server model. Skype has a free on-net VoIP service (SkypeIn), and a fee-based off-net service (SkypeOut) that allows calls to PSTN and mobile phones.

JAJAH [9] and Gizmo Project [10] are other examples of peer-to-peer VoIP network systems.

Skinny Client Control Protocol (SCCP) [11] is a Cisco proprietary protocol for real-time calls and conferencing over IP. It is used between Cisco Call Manager (H.323 proxy) and Cisco VoIP phones (Skinny client).

MiNET [12] is a proprietary VoIP protocol used by Mitel phones and PBXs.

Siemens developed CorNet IP [13], a protocol that allows the communication and interworking in line-switched, packet-switched and hybrid network environments.

The Inter-Asterisk eXchange (IAX) protocol [14] provides control and transmission of streaming media over IP networks. It is used by the Asterisk VOIP PBX as an alternative to SIP and H.323 to connect to other devices that support IAX.

3. Voice-over-IP Protocols Comparative Study

Although this paper presents the four more relevant standard VoIP protocols, we will further examine only the more widespread ones: H.323 and SIP. After above-mentioned capabilities, architecture and protocol stack, we complement this study with a summary shown in Table 1. It presents a comparison based on the following characteristics: client type, network intelligence and services, model used, addressing mechanisms, message definition and encoding, media transport, transport protocol, authentication and encryption, capability negotiation, PSTN internetworking, support for video and data conferencing, and available open-source projects [4, 15].

H.323 and SIP have some common characteristics like the use of intelligent endpoints, and the employ of the same media transport and transport protocols. However they differ on the other analyzed characteristics that partially support the reason of H.323 success. For instance, H.323 model is based on telephony while SIP's model is based on Internet/WWW. H.323 has flexible addressing mechanisms, in opposition to SIP that only supports URI addresses. H.323 uses the standardized notation Abstract Syntax Notation One (ASN.1) for message definition, and encodes messages in binary format. SIP defines messages on Augmented Backus-Naur Form (ABNF) and encodes them in ASCII text format, leading to large messages with consequences for bandwidth, delay and processing [4]. Authentication and encryption is available on both of them through the use of different protocols. In terms of capability negotiation SIP presents itself more limited than H.323. Regarding PSTN internetworking, H.323 uses some traditional PSTN protocols like Q.931 that enables PSTN integration. SIP has no commonality with PSTN [4]. Considering video and data conferencing SIP also presents limited or no support in opposition to H.323. Finally, both protocols have open-source projects.

Table 1. Comparison between H.323 and SIP characteristics

	H.323	SIP
Client Type	Intelligent	Intelligent
Network Intelligence and Services	Provided by Gatekeepers	Provided by Servers (Proxy, Redirect, Registrar)
Model Used	Telephony/Q.SIG	Internet/WWW
Addressing	E.164, URI, E-mail address	URI
Message Definition and Encoding	ASN.1 - Binary	ABNF - ASCII
Media Transport	RTP/RTCP, SRTP	RTP/RTCP, SRTP
Transport Protocol	TCP/UDP	TCP/UDP
Authentication and Encryption	H.235	HTTP (Digest and Basic) - SSL, PGP, S/MIME
Capability Negotiation	Good	Limited
PSTN Integration	Well suited	Non-native
Video Conferencing	Full support	Limited support
Data Conferencing	Full support	No support
Open-Source Projects	H.323 Plus [16]	reSIProcate [17]

Both H.323 and SIP were introduced in 1996 and are considered second generation protocols. These protocols primarily aimed at delivering voice and video service only, and adding new functionalities is quite complicated since they exhibit a monolithic approach. Furthermore, they did not take the full potential of the IP network, and QoS, security, NAT and Firewall traversal issues were not considered from the ground up.

The Advanced Multimedia System (AMS) is a new multimedia system project driven by the ITU-T SG16 [18], that is viewed as the successor to the legacy H.323 and SIP.

According to [4], the AMS project intends to develop a third generation multimedia terminal and system architectures that supports distributed and media rich collaboration environments, and that will expand the capabilities available in existing multimedia systems. It will enable new applications with minimum or no changes to deployed infrastructure, and enable third-party application developers to add new functionality to the system. Furthermore, issues like QoS, security, NAT and Firewall traversal will be addressed from the beginning.

The AMS project description [19] has been presented on September 2007, and SG16 is now collecting requirements while considering possible architectures for the system.

4. Conclusion

In this paper, we started with an introduction to the VoIP technology, providing some insight on the most important standard VoIP protocols. After, the two more widespread standards (ITU-T's H.323 and IETF's SIP) were compared. H.323 solid foundation and technical capabilities are reasons for its success over SIP. However, SIP is less complex than H.323. Therefore, nowadays, SIP is increasing its utilization whereas H.323, given its complexity, is being less used.

However, the advancement of technology and the advent of the Next Generation Networks, together with the technical limitations of these protocols create the opportunity for the introduction of novel approaches. AMS aims to take advantage of emerging technologies, advances in proven technologies, and deeper understanding of existing ones, to create new and improved forms of communication capabilities.

Acknowledgment

Part of this work has been supported by the Group of Networks and Multimedia of the Institute of Telecommunications – Covilhã Lab, Portugal.

References

[1] D. Cohen, "Specifications for the Network Voice Protocol (NVP)", RFC 741, November 1977, ftp://ftp.rfc-editor.org/in-notes/rfc741.txt

[2] ITU-T, "ITU-T Recommendation H.323 (06/2006)", Packet-based multimedia communications systems, June 2006, http://www.itu.int/rec/dologin_pub.asp?lang=e&id=T-REC-H.323-200606-I!!PDF-E&type=items

[3] O. Hersent, J.-P. Petit, and D. Gurle, *IP Telephony - Deploying Voice-over-IP Protocols*: John Wiley & Sons, Ltd, 2005.

[4] Packetizer Inc, "Packetizer", http://www.packetizer.com/, accessed at January, 2008.

[5] "H.323 Forum", http://www.h323forum.org/, accessed at January, 2008.

[6] J. Rosenberg, H. Schulzrinne, G. Camarillo, A. Johnston, J. Peterson, R. Sparks, M. Handley, and E. Schooler, "SIP: Session Initiation Protocol", RFC 3261, June 2002, ftp://ftp.rfc-editor.org/in-notes/rfc3261.txt

[7] B. Foster and C. Sivachelvan, "Media Gateway Control Protocol (MGCP) Return Code Usage", RFC 3661, December 2003, ftp://ftp.rfc-editor.org/in-notes/rfc3661.txt

[8] Skype, "Skype official website", http://www.skype.com/, accessed at January, 2008.

[9] JAJAH, "JAJAH - web-activated telephony", http://www.jajah.com/, accessed at January, 2008.

[10] Gizmo Project, "Gizmo - A free phone for your computer", http://www.gizmoproject.com/, accessed at January, 2008.

[11] Cisco Systems, "Cisco Unified Communications Manager", http://www.cisco.com/en/US/products/sw/voicesw/ps556/products_tech_note09186a0080624977.shtml, accessed at January, 2008.

[12] Mitel Networks Corporation, http://www.mitel.com/jsp/index.jsp, accessed at January, 2008.

[13] Siemens, http://enterprise.usa.siemens.com/products/products/infrastructure/cornet.html, accessed at January, 2008.

[14] M. Spencer and F. W. Miller, "IAX Protocol Description", 2004, http://www.voipinfo.hu/content/download/iax.pdf

[15] Cisco Systems Inc., "Overview of the Session Initiation Protocol", http://www.cisco.com/univercd/cc/td/doc/product/voice/sipsols/biggulp/bgsipov.htm#xtocid8, accessed at January, 2008.

[16] Packetizer Inc., "H.323 Plus, The Standard in Open Source H.323", http://www.h323plus.org/, accessed at February, 2008.

[17] reSIProcate, "Home of the reSIProcate projects", http://www.resiprocate.org/Main_Page, accessed at February, 2008.

[18] ITU-T Study Group 16, "Advanced Multimedia System (AMS) - Third generation of ITU-T Multimedia Systems and Terminals", http://www.itu.int/ITU-T/studygroups/com16/ams/index.html, accessed at January, 2008.

[19] ITU-T, "Advanced Multimedia System (AMS) Project Description", 2007, http://www.packetizer.com/ipmc/ams/papers/ams_project_description.pdf