Coming issue: "Present and Future of the Informatics Profession"

Voice over IP
Guest Editors: David Fernández Cambronero and Eberhard Zangger

Joint issue with NOVÁTICA and INFORMATIK/INFORMATIQUE

Signalling in Voice over IP Networks

José Ignacio Moreno, Ignacio Soto and David Larrabeiti

Voice signalling protocols have evolved, keeping with the prevalent move from circuit to packet switched networks. Standardization bodies have provided solutions for carrying voice traffic over packet networks while the main manufacturers are already providing products in workgroup, enterprise, or operator portfolio. This trend will accrue in next years due to the evolution of UMTS mobile networks to an "all-IP" environment. In this paper we present the various architectures that are proposed for signalling in VoIP, mainly: H.323, SIP and MGCP. We also include a brief summary about signalling in classical telephone networks and, at the end, we give some ideas about the proposed "all-IP" architectures in UMTS 3G mobile networks.

1 Signalling in telephone networks

Signalling in classic telephone networks has evolved dramatically during the 20th century, at the pace driven by the development of circuit switching technologies these networks are based on. The manual switches in use at the end of the 19th century were replaced by electro-mechanical switching in the advent of the 20th century. This technological stage would last until the 60s. Signalling was carried "in-band" (level change and tones inside the telephone channel), and was interpreted by electromechanical and electronic elements (relays and filters) on its way through the network. End-to-end physical connectivity quickly evolved to logical connectivity, and transmission, formerly analog FDM, moved to digital TDM structured as 64 Kb/s channels.

In the middle 60s, transmission and switching integrated digital network merged with the advent of digital exchanges and CPU-controlled switching (Stored Program Control concept). The 64 Kb/s synchronous channels are byte-by-byte switched in space and in time. These exchanges are integrally controlled by processors that exchanges a signalling protocol with other exchanges' processors. The first signalling protocols used in these systems were based on the status of a few bits in the TDM frame permanently attached to each voice channel, just as binary representations of predecessor analog systems. The quantum leap in the history of telephone signalling was the application of computer networks technology to the design of the signalling system. Signals became messages exchanged by systems over an independent packet switching network exclusively dedicated to this task.

Although this type of operation is now almost ubiquitous in the public telephone network, the last segment to be digitised – the subscriber loop – remains analog, with a small deployment of fully digital accesses (ISDN). As a result, existing user-network signalling has evolved very little as compared to the revolution observed in the network-to-network architecture, that has enabled the development of many additional services, mobile telephony, intelligent network services, broadband ISDN and internetworking with VoIP systems.

The network-to-network signalling system that has supported this evolution is the *Signalling System number 7* (SS7). The first CCITT standard is dated 1981 ("Yellow Book"), and has been refined and enhanced in successive editions in 1985 ("Red Book"), in 1989 ("Blue Book") [SS7 89] and following ITU-T standards.

SS7 is a complete protocol stack where signalling units are messages issued by signalling applications transported in packets. The essential features of this system are:

- Bundles of signalling links and nodes build a packet switching network that is logically independent from the circuit switching one, with a specific numbering plan and internationally defined by ITU-T.
- SS7 is a common channel signalling system. A set of channels between Signalling Points at exchanges (and Signalling Transfer Points) is dedicated to transport signalling to setup, release, supervise, etc. any voice or data 64 Kb/s channel. In previous signalling systems, a signalling associated to each voice circuit was transported over a transmission channel exclusively dedicated to it.
- SS7 is a 4-level protocol stack (Figure 1).

Jose Ignacio Moreno is Assistant Professor of Computer Networking at *Universidad Carlos III* of Madrid. He obtained his Ph.D. at the Technical University of Madrid in 1996. From 1991, he has participated in several national and international research projects in the field of telematic services and networks. <jmoreno@it.uc3m.es>

Ignacio Soto is Associate Professor of Computer Networking at *Universidad Carlos III* of Madrid. He obtained his Ph.D. at Vigo University and worked in different National and European research projects. <isoto@it.uc3m.es>

David Larrabeiti is Assistant Professor of Computer Networking at *Universidad Carlos III* of Madrid. He obtained his Ph.D. at the Technical University of Madrid in 1996. From 1991, he has participated in several European research projects related to the study and deployment of next generation networks and services. <dlarra@it.uc3m.es>
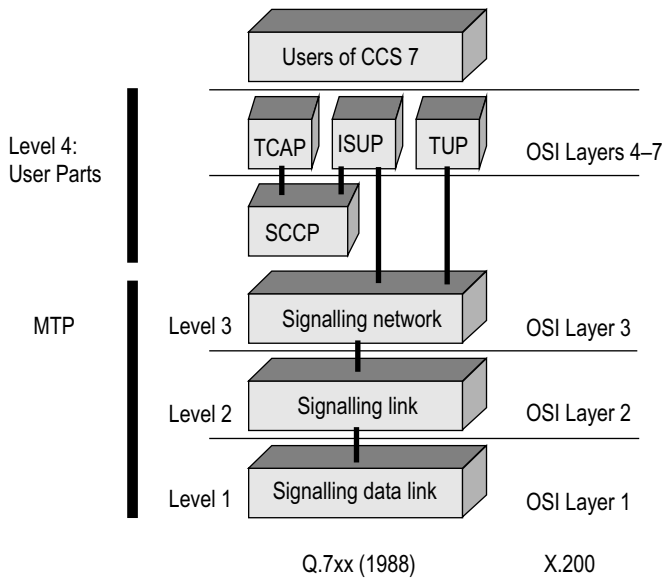
Fig. 1: The SS7 protocol stack from the OSI perspective

Telephony signalling networks have been specifically designed to – and implicitly constrained to – operate and manage 64 Kb/s circuit switched channels. There is now a strong tendency to move towards the use of packet switched networks. It is interesting to observe how packet switching, introduced in legacy networks to add flexibility and reliability to signalling in the control plane of protocol stacks, is being extended today to the user plane to carry packetized voice, merging again control and voice traffic.

2 Videoconference over packet Networks: H.323

ITU-T was the first standardization body which created an standard for transferring multimedia traffic over IP networks. The standard H.323 version 1 appeared on 1996 and was called "Visual telephony systems and terminal equipment for local area networks which provide a non-guaranteed quality of service". As a result, H.323 started the way to provide multimedia –and therefore voice transmission– over packet networks. The main contribution of H.323 was the provision of signalling protocols for controlling all of these communications, as media transmission architecture (voice, video, data) was adopted from previous works of IETF (mainly RTP/RTCP [RFC 1889] protocols tested on MBONE initiative).

After this initial version, on 1998 H.323v2 appeared with a new name, "Packet based multimedia communications systems", which have remained until now (version 4 was approved on Nov. 2000 [ITU 00a]). H.323 is considered an umbrella of standards and consists of 4 types of functional elements (Figure 2):

Figure 2: H.323 Architectural Model

- *H.323 Terminal*, is an endpoint on the network which provides for real-time, two-way communications with another H.323 terminal, Gateway, or Multipoint Control Unit. A terminal may provide speech only, speech and data, speech and video, or speech, data and video. The block functional structure of an H.323 terminal is shown in the Figure 3.

Fig. 2: H.323 Architectural Model

Fig. 3: H.323 Terminal structure

- *H.323 Gateway (GW)*, is an endpoint on the H.323 network which provides for real time, two-way communications between H.323 Terminals and other ITU Terminals connected, for example, to PSTN, ISDN, broadband ISDN (ATM), or QoS enhanced LANs. The purpose of the Gateway is to reflect the characteristics of a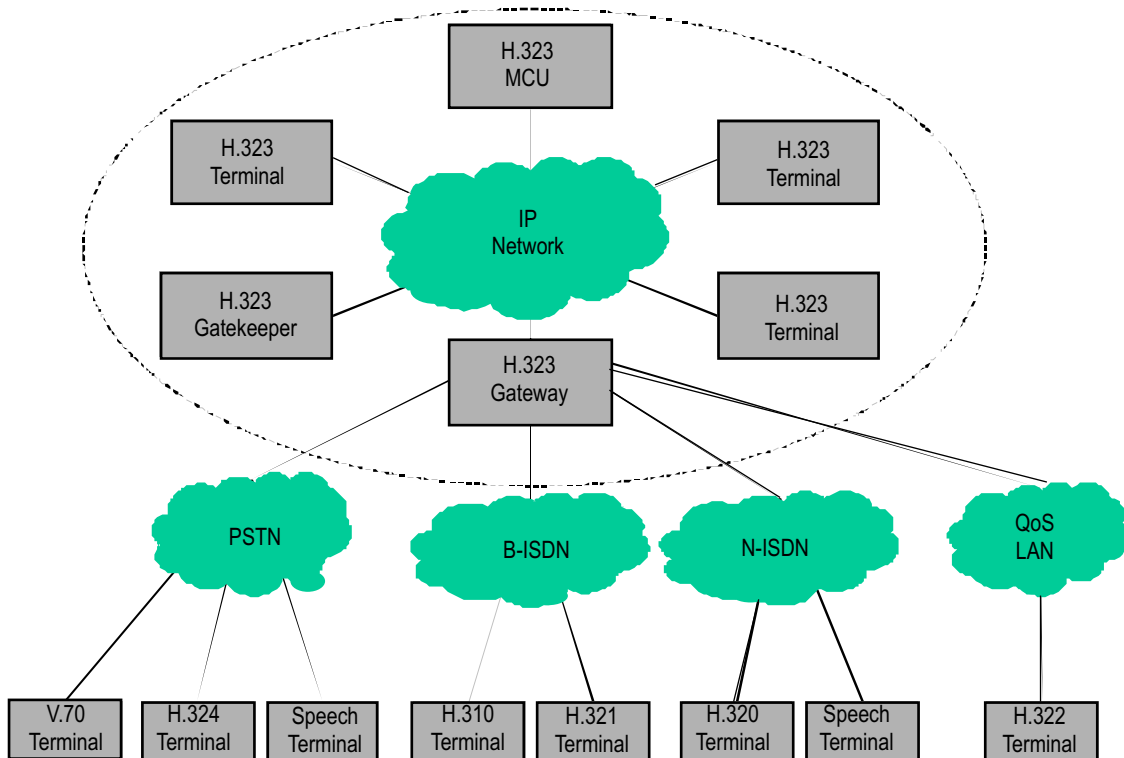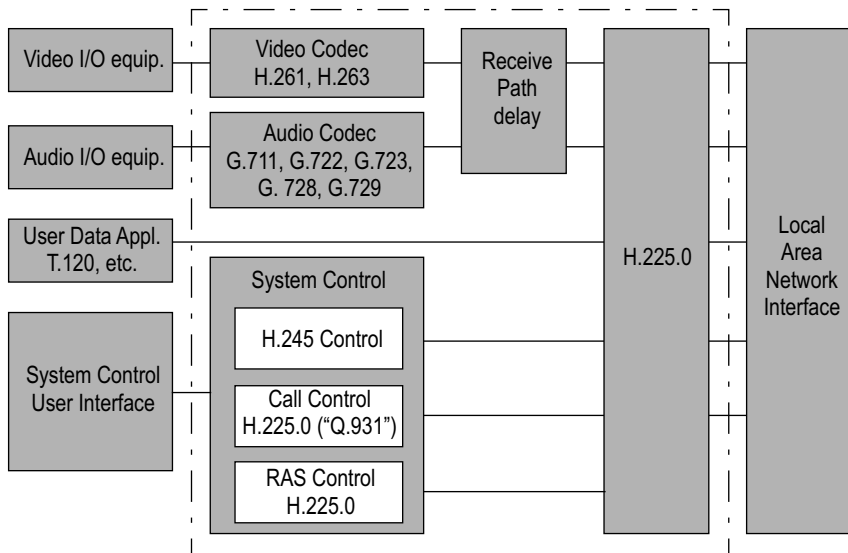 network endpoint to a Switch Circuits Network (SCN) endpoint, and the reverse, in a transparent way. Functions like translation between transmission formats and between communications procedures must be provided.
- *Multipoint Control Unit (MCU)*, is an endpoint on the network which provides the capability for three or more terminals and Gateways to participate in a multipoint conference. It may also connect two terminals in a point-to-point conference which may be later developed into a multipoint conference. The MCU consists of two parts: a mandatory Multipoint Controller (MC) which provides capability negotiation between all terminals to achieve common levels of communications and optional Multipoint Processors (MP), which performs mixing or switching of audio, video and data. This functionality could be integrated in an H.323 terminal.
- *Gatekeeper (GK)*, is an H.323 entity on the network that provides call control services to the H.323 endpoints. This element is the key block of the H.323 architecture for development of services and for the application of this technology in a medium-large environment. In principle the GK is an optional block of the architecture that was decided in order to boost the development of H.323 terminals without the requirement of a complex element, but without this element the model is quite limited. GK provides important services like: address translation and directory services,

admission and bandwidth control, call control signalling or call authorization.

The complete protocol stack used in H.323 is shown in the Figure 4, including media and signalling transmission planes. Most of the control channels use TCP as transport protocol, but from version 3 on, UDP can also be used.

H.323 entities establish connections in different phases. Considering a normal scenario, an H.323 session between two entities in the same zone (that is, depending on the same GK), will involve the following steps (Figure 5):

1. *Phase A: Call Setup.* The Calling entity uses RAS (Registration Admission and Status) protocol to send an Admission Request Message (ARQ) to the GK, requesting authorization to make a call and providing the identification of the called party. The GK can accept the call by sending a confirmation (ACF message) or reject it (ARJ message). When the call is accepted, the calling party establishes a new TCP connection to the address provided by the GK in the ACF message. This connection uses H.225.0 protocol to send a Setup Q.931 message. The Called party must first connect to the GK and ask for permission to accept the connection. When granted, the called party sends a Connect message, including the H.245 Control Channel Transport Address for use in the next phase.

2. *Phase B: Initial communication and capability exchange (H.245).* In this phase, a new TCP connection is established between calling and called party to exchange ability and information about media channels using H.245 protocol. After this step, both parties have agreed on the media parameters (codecs, samples per frame, etc.) and exchanged information about media channels (ports, etc.). This TCP connection must remain until call termination phase and is used to open or close media channels or modify their parameters.

3. *Phase C: Establishment of audiovisual communication.* In this step, both entities exchange multimedia information directly using a RTP/UDP/IP for the media channels. They
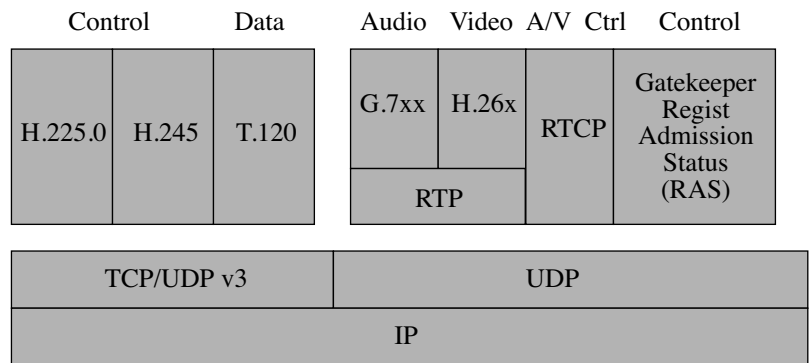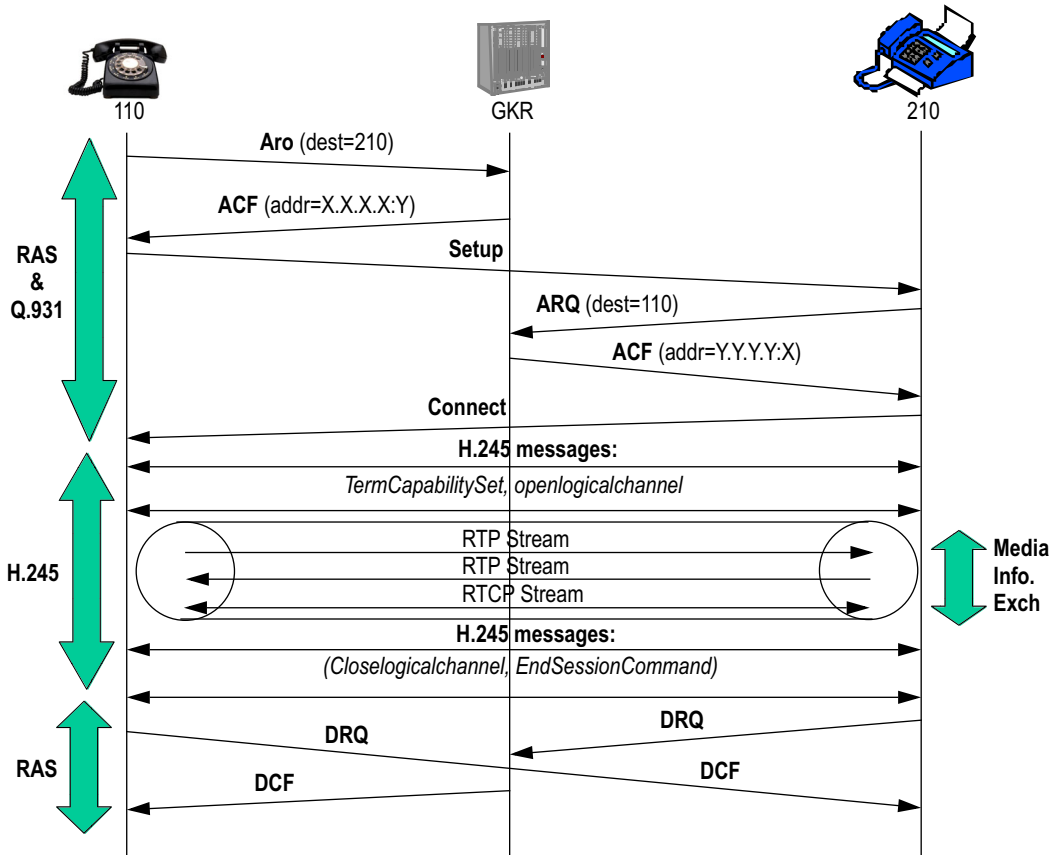

Fig 4: H.323 protocol stack

Fig. 5: Generic Call Flow

also use RTCP messages as a control channel to get feedback information about how their flows are being received.

4. *Phase D: Call termination*. H.323 entities must inform each other about the call termination by sending an *EndSession-Command* message, which will produce the closing of H.245 channel. Besides, they must send a *Disengage Request* (DRQ) to the GK, by using RAS channel, in order to inform the GK about the termination of the call. GK then performs actions such as release resources, provide billing information, etc.

The previous example shows that H.323 requires a lot of connections before terminals can exchange information. This is one of the main drawbacks of H.323v1 that was partially solved in version 2 using two possible optional modes: Fast Connect Procedure, which allow to open media channels during H.225.0 phase; and H.245 tunnelling which uses the same channel for transmitting H.225.0 and H.245 messages.

3   IETF's proposal for VoIP: SIP

The Session Initiation Protocol (SIP) is an application protocol, defined in RFC2543 [RFC 2543], that is being designed by the IETF MMUSIC (Multiparty Multimedia Session Control) working group to enable users to participate in multimedia sessions, that is, to establish, modify and terminate multimedia sessions calls. MMUSIC working group [MMUSIC] focus on loosely coupled conferences as they exists today on the MBONE. One of the main issues in this area is related with how to inform users about forthcoming sessions, media requirements, addresses, etc. There are two basic ways to locate and to participate in a multimedia session:

- *Advertisement*. Sessions are advertised in various ways like e-mail, web pages, newsgroups or a multicast advertisements via Sessions Announcement Protocol (SAP) like in the MBONE.
- *Invitation*. Users are invited by others to participate by using the Session Initiation Protocol (SIP).

SIP has been proposed as a generic unicast and multicast initiation protocol and also as an IP Telephony call set-up protocol. It is based on a client-server protocol. SIP clients send a *Request Message* for a service, and a server handles the request, answering with a *Response Message*. SIP terminals can both generate and receive request as they are composed by a User Agent Client (UAC) and a User Agent Server (UAS).

SIP terminals can establish voice calls directly without requiring any other element. Figure 6 shows an example in which user1 calls user2 by sending an INVITE Request primitive containing user1 supported capabilities for receiving audio and a UDP port (port 12345 and mlaw codec). When user2 receives INIVITE Request, he can establish a voice channel to 12345 UDP port of user1 while he accepts the request by sending an OK Response message. In addition, user2 response includes its own media capabilities, which are used by user1 to establish a voice channel (GSM codec at 54321 port in the example) and send an ACK message to acknowledge user2's response. In order to terminate the connection, any of the
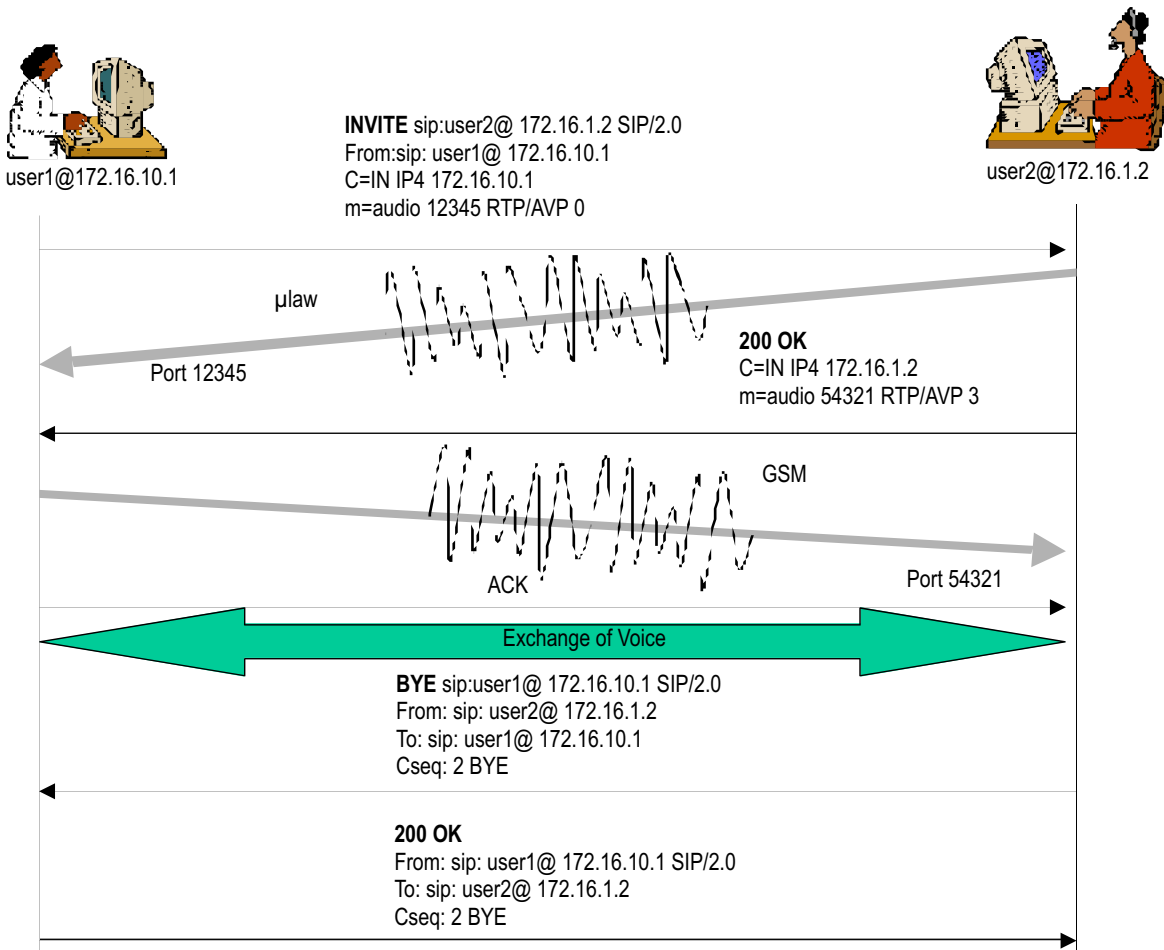
INVITE sip:user2@ 172.16.1.2 SIP/2.0
From:sip: user1@ 172.16.10.1
C=IN IP4 172.16.10.1
m=audio 12345 RTP/AVP 0

user1@172.16.10.1

user2@172.16.1.2

μlaw

Port 12345

200 OK
C=IN IP4 172.16.1.2
m=audio 54321 RTP/AVP 3

GSM

ACK

Port 54321

Exchange of Voice

BYE sip:user1@ 172.16.10.1 SIP/2.0
From: sip: user2@ 172.16.1.2
To: sip: user1@ 172.16.10.1
Cseq: 2 BYE

200 OK
From: sip: user1@ 172.16.10.1 SIP/2.0
To: sip: user2@ 172.16.1.2
Cseq: 2 BYE

Fig 6:
Simple SIP call

parties can send a BYE Message, which must be acknowledged by an OK response.

SIP messages are encoded using HTTP/1.1 message syntax [RFC 2068], and the content of each message follows session description protocol (SDP) [RFC 2327], heavily used in the context of MBONE for distributing information about MBONE sessions.

In addition to SIP terminals which represent an IP telephones or Gateways, SIP architecture is based on four different servers entities:

• *Proxy server,* which forwards requests to its final destination. Like the Gatekeeper in H.323 model, it receives requests and sends them to the appropriate destination. In this case the Via field in the request/response message indicates the intermediate proxies which participate in the request process. This avoids routing loops as well as permits to force the responses to follow the same route back. Proxy servers do not need to relay the media except in case of transcoding operations (they only need to relay control information).

• *Redirect Server,* which, on the contrary to proxies, does not forward INVITE requests but responds to clients with a redirection message that informs about the next hop.

• *Registrar server,* that keeps track of the current location of a user by accepting registration request messages. Registrar

server are discovered by using well-known multicast addresses or preconfigured unicast addresses.

• *Call Agent,* which is a service that handles incoming and outgoing calls on behalf of a user. It is a mixture of a proxy, a registrar and a redirect server. A call agent can perform tasks like:
  • Try to find a user by redirecting the call setup message to the proper location or locations.
  • Implement call redirection rules such as call forwarding on busy, call forward on no answer, etc.
  • Implement call filtering with origin/time-dependant rules.
  • Record unsuccessful call attempts for future reference.
  • Perform any other call management task on behalf of the user.

The objects addressed by SIP are users at hosts or servers identified by Uniform Resource Identifiers [RFC 2396]. The SIP URI has the form: "user@host", where "user" is a user name or a telephone number and "host" is a domain name or a numeric network address.

Figure 7 shows a complete example of an interaction between SIP servers. David from company.es wants to call jmoreno@upm.es. So, he sends a request to his SIP server (sip.company.es), which acts as a proxy and relies the *INVITE* request to the SIP server of upm.es (sip.upm.es). This server,
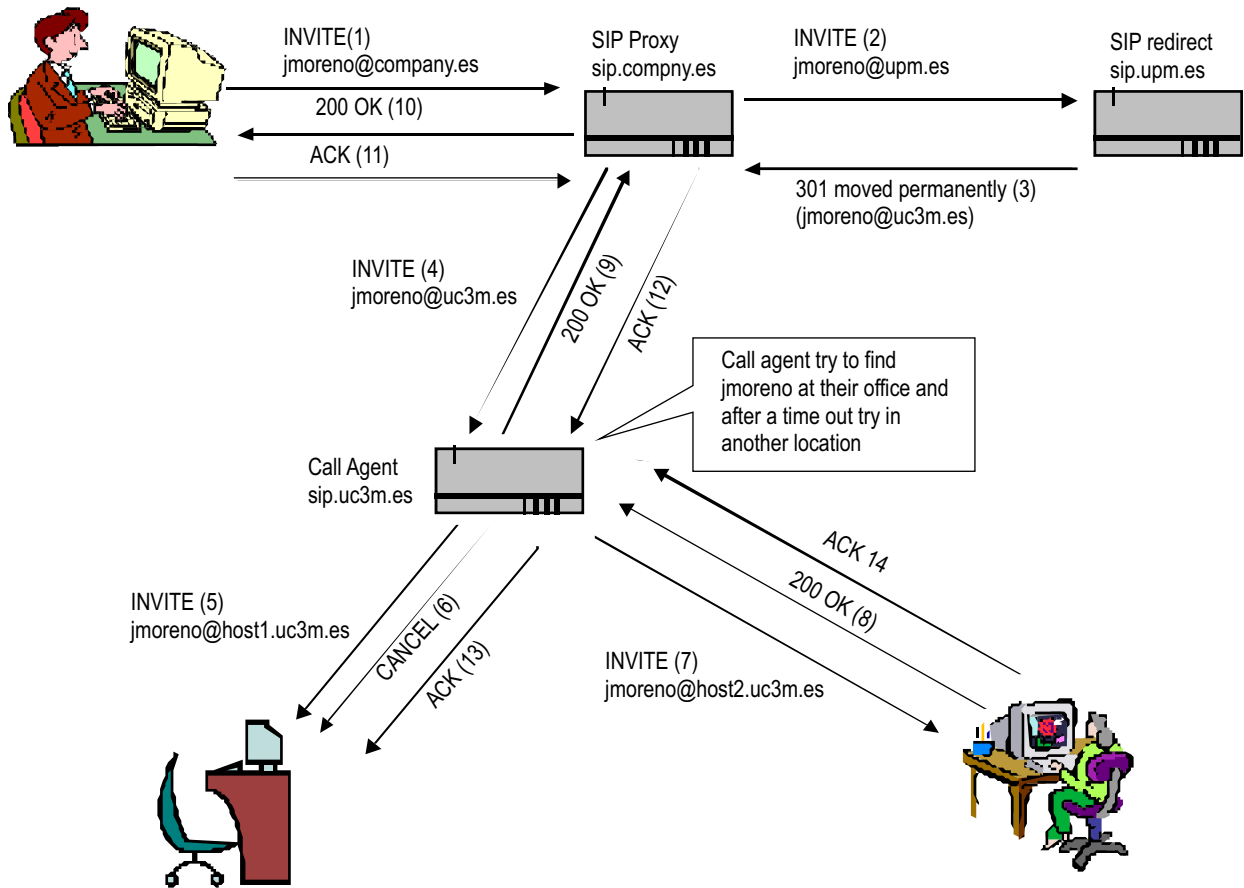
Fig. 7: Example of SIP servers

which acts as a redirect server, redirects the call by answering with a moved permanent pointing to the new address jmoreno@uc3m.es. SIP server at company.es progresses the call contacting SIP server at uc3m.es. This server, which act as Call agent, relays the request to the usual location of jmoreno at host1.uc3m.es, but after a timeout without answer from jmoreno, *CANCELs* the previous request and redirects the call to another usual location of this user in a different branch office (host2.uc3m.es). The user answers the call by sending a *200 response* which is sent back to the calling party.

One of the main advantages of SIP is its simplicity, which translates into simpler implementations and shorter delays. While H.323v1 needs 4 or 5 round-trips to establish a connexion, SIP requires only one. As mentioned before, this key aspect was corrected in H.323 from version 2 on.

4   VoIP in the core: MEGACO and MGCP

H.323 and SIP were developed bearing in mind terminals connected directly to IP networks. Both of them consider connections between IP terminals or between IP terminals and conventional SCN terminals by means of Gateways. The goal of MEGACO architecture is, in addition, to establish connections between SCN terminals through IP-based networks. The main idea comes from operators, especially incomers: What about deploying IP-based core infrastructure for new invest-

ments in the network while providing the same telephone service?

MEGACO solves this problem mainly by splitting Gateways into three different entities:

* *Media Gateway Controller (MGC)*, which provides the H.323 or SIP signalling and performs mapping between SCN signalling protocols and IP-based signalling protocols.
* *Media Gateway (MC)*, which provides media mapping and transcoding functions. It terminates SCN (PCM signal typically) and packet media signals and performs address translation, echo cancellation, playing announcements, receiving/sending DTMF digits, etc.
* *Signalling Gateway (SG)*, which provides signalling mediation between IP and SCN domains.

In a common scenario, these three elements are intended to be physically separated, providing advantages like, concentration of many MG in a few MGC controlled by a SG. Figure 1 shows MEGACO architecture.

Media Gateway Control Protocol (MGCP) provides a simple client/server model to control Media Gateways. MGCP is the result of previous protocols and has been proposed to different standardization groups like MEGACO IETF working group [RFC 2705] [RFC 30115] and ITU-T [ITU 00b]. MGCP uses SDP to exchange parameters to the gateway (IP, UDP port, media, etc.).
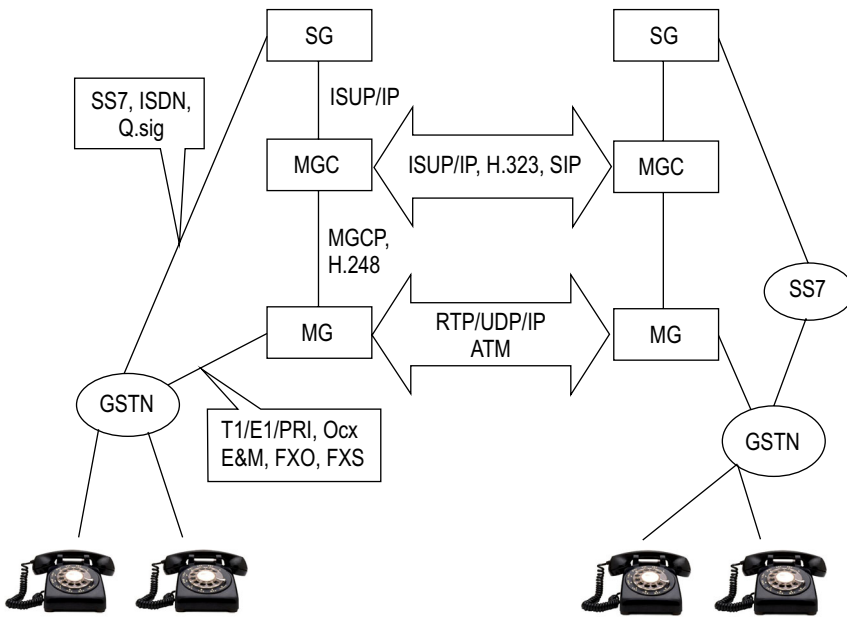
Fig 8: Megaco Architecture

5 Third Generation Mobile Networks: Towards and All-IP architecture

The Third Generation Partnership Project (3GPP) [3GPP 00a] works in the standardization of a 3G mobile system based on an evolved GSM core network and WCDMA radio access technologies. This system is the Universal Mobile Telecommunication System (UMTS) that will be launched to the market in the near future.

The first phase of the specification of UMTS was finished at the beginning of year 2000 and was called Release 1999 (R99). The 3GPP continues developing specifications to set the path of the evolution of UMTS systems. Release 4 and Release 5 (with target dates as of December 2001) are the following stages on this evolution. In this section we briefly describe the UMTS R99 architecture and the planned evolution for this architecture towards the inclusion of VoIP protocols in it [3GPP 00b].

As shown in Figure 9, the UMTS R99 architecture [3GPP 00c is basically a GSM/GPRS network [Bettstetter et al. 99 with a new access network named UTRAN (UMTS Terrestrial Radio Access Network). UTRAN is made of Radio Network Controllers (RNC) and B Nodes, that can coexist with classical GSM access network nodes (BTS and BSC). The architecture is clearly divided in two parts: *Circuit Switched (CS)* domain, used to carry voice traffic (made of MSC and GMSC nodes); and *Packet Switched (PS)* domain, used to transport data traffic
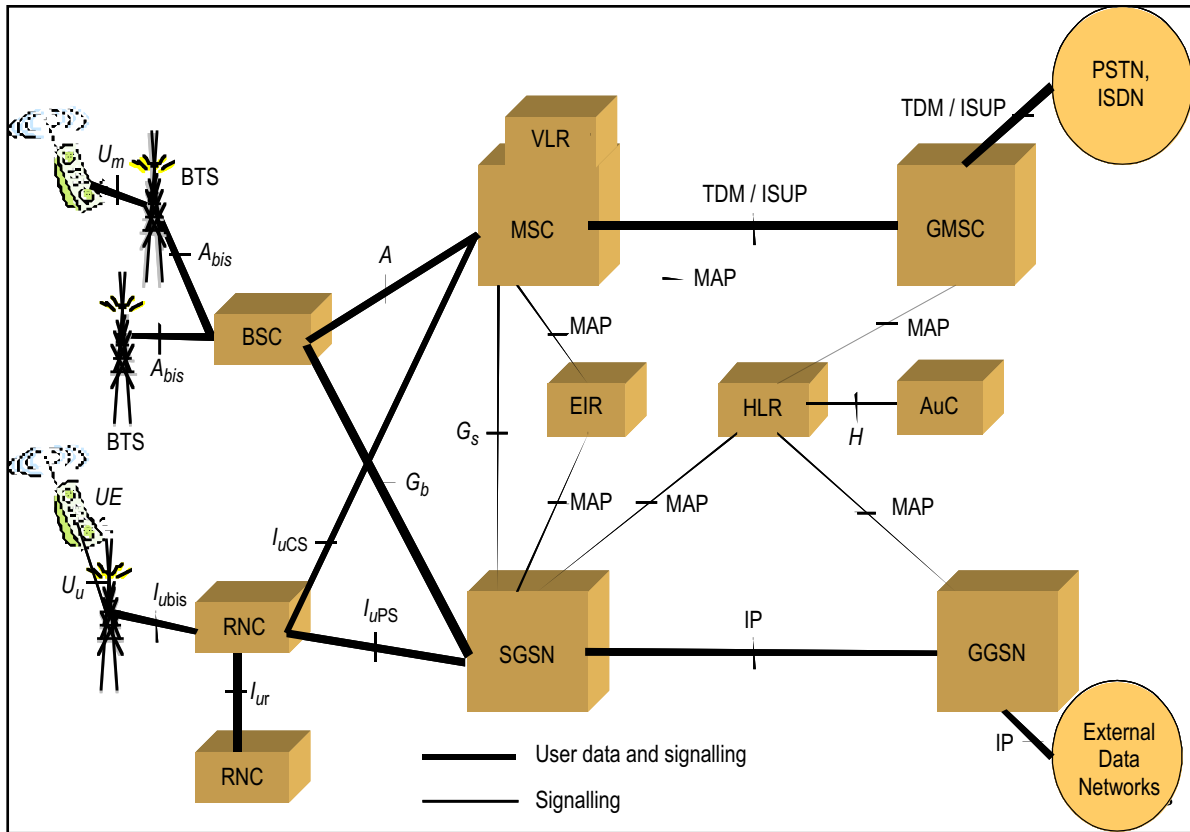

Fig. 9: UMTS R99 Architecture

(made of SGSN and GGSN nodes). Other nodes in the architecture are used to keep information about users (VLR, HLR, EIR, AuC).

The main advantage of this design is that allows an easy evolution towards UMTS systems starting from existing GSM/GPRS networks. However, there is a clear tendency to unify both domains using an IP based core network, as Release 4 and 5 point out. Even, IP could be used as the packet switching technology in the access network – not only in the core – as different European projects are proposing [MobyDick]. All this proposals have been globally named the "all-IP architecture" [Bos/Leroy 01]. The reason behind this tendency is that packet switching networks are cheaper, efficient and capable of carrying all the different classes of traffic. Besides, IP is a proved protocol that allows a seamless intercommunication with Internet.

Release 4 and 5 of UMTS specifications define the evolution steps towards the "all-IP" architecture. Key points on this path are:

- The CS domain, which was originally designed to use "classical" TDM technologies, is evolved to make it independent of the transport technology, allowing it to work over IP and ATM core networks. This evolution is highly based on the ideas developed in MEGACO architecture. For example, MSC nodes are divided in two elements, the MSC server responsible for all signalling and control functions, and the Media Gateway (MGW) responsible of data transport functions. It also uses other VoIP protocols like H.245 or RTP.
- A new subsystem, named "IP multimedia" (IM), is added to the PS domain, in order to support IP multimedia services based on SIP. The central entity in IM is the Call State Control Function (CSCF), which is in charge of call set-up and termination, routing of incoming calls, address handling, etc. It basically includes much of the functionality found on SIP servers (proxies, registrars, etc.). One of the main contributions of IM subsystem will be the use of "SIP phones", that is, mobile terminals that directly support SIP signalling.

6   Conclusions

Signalling in telephone networks is clearly evolving from circuit switch SS7 based networks to IP centric solutions. A rich set of standardized and proprietary solutions have appeared in recent years to carry voice traffic over IP networks and cope with problems like addressing, admission and call control, internetworking with existing networks, etc. Work has concentrated over two main scenarios: voice terminals directly connected through IP networks, and operators that use IP backbones to connect traditional switched circuit networks. The first scenario can be solved by means of H.323 or SIP proposals, and the later by MEGACO or H.248.

Today, different operators and enterprises are using or beginning to use these technologies to carry voice traffic over packet networks. The trend will be increased in the near future with the evolution of mobile UMTS networks towards an "all-IP" technology, in which multimedia services will be deployed. As a result, in the future voice traffic will be mainly carried over IP technology. However, the path towards "all-IP" networks will not be easy, there is still a lot of problems to be solved, most of them related to interoperability with existing networks, and it will take a long time, investment in traditional networks has to be recovered.

7. References

[SS7 89]
  Specifications of Signalling System Nº 7. CCITT Blue Book, fascicle VI.7, recommendations Q.701-Q.716, Q.721-Q.766, Q.771-Q.795. ITU, 1989.
[RFC 1889]
  RFC 1889. H. Shulzrinne, S. Castner, R. Frederick, V. Jacobson. RTP: A transport protocol for real time protocol.
[ITU 00a]
  ITU-T Recommendation H.323: "Packet-based Multimedia Communications Systems", November 2000
[RFC 2543]
  RFC 2543. M. Handley, H. Shulzrinne, E. Schooler, E. Rosenberg. SIP: Session Initiation Protocol.
[MMUSIC]
  MMUSIC web site. http://www.ietf.org/mmusic
[RFC 2068]
  RFC 2068. R. Fielding and others. Hypertext Transfer Protocol - – HTTP/1.1
[RFC 2327]
  RFC 2327. 2327 M. Handley, V. Jacobson. SDP: Session Description Protocol.
[RFC 2396]
  RFC 2396. T. Berners-Lee, R. Fielding, Uniform Resource Identifiers (URI): generic syntax.
[RFC 2705]
  RFC 2705. M. Arango et al., "Media Gateway Control Protocol (MGCP)".
[RFC 30115]
  RFC 3015. F. Cuervo, N. Greene, A. Rayhan et al., "Megaco Protocol Version 1.0"
[ITU 00b]
  ITU-T Recommendation H.248: "Gateway Control Protocol", June 2000
[3GPP 00a]
  3GPP web site: http://www.3gpp.org.
[3GPP 00b]
  3GPP Technical Specification TS 23.002, v5.0.0: Network Architecture (Release 5). October, 2000.
[3GPP 00c
  3GPP Technical Specification TS 23.002, v3.3.0: Network Architecture (Release 1999). March, 2000.
[Bettstetter et al. 99
  C. Bettstetter, H.-J. Vögel, J. Eberspächer; GSM Phase 2+, General Packet Radio Service GPRS: Architecture, Protocols, and Air Interface; IEEE Communications Surveys Vol. 2, No. 3, 1999.
[MobyDick]
  MobyDick Project. http://www-int.berkom.de/mobydick
[Bos/Leroy 01]
  Lieve Bos, Suresh Leroy; Toward an All-IP-Based UMTS System Architecture; IEEE Network, Vol. 15, No. 1; 2001.