

Packet Based Telephony

Boštjan Vlaovič, Zmago Brezočnik

Faculty of Electrical Engineering and Computer Science, University of Maribor

Smetanova ul. 17, SI-2000 Maribor, Slovenia

e-mail: {bostjan.vlaovic,brezocnik}@uni-mb.si

Abstract

This paper describes methods used to place telephone calls over packet-based networks with emphasis on Internet point-to-point communication. Two approaches will be described — H.323 and SIP. The first is ITU standard and is already widely deployed by commercial vendors. The second, SIP, was developed by IETF activists and represents “Internet approach” to telephony.

First quick overview of reasons for growing importance of IP telephony will be given. It will be followed by quick introduction to voice transmission over packet networks and overview of the existing standards. In the second part of the paper short example of H.323 and SIP call setup will be given.

1 Introduction

Since the initial deployment of the Internet, the architecture of the network has been changing. It has evolved in response to technological progress, experimentation, explosive growth, and new services. For quite some time a dream of many developers and users is to use multimedia applications over the Internet. One of the first objectives, to transfer voice over Internet Protocol (IP) in real-time, is already available in many commercial and freely available products. First successful VoIP (Voice over Internet Protocol) connection was made by Informational Sciences Institute (University of Southern California) and Lincoln’s Laboratory (Massachusetts Institute of Technology) in the late 1974. In 1976 first RFC (Request For Comment), RFC 741 (Specifications for the network voice protocol), was published. Due to various reasons work on VoIP stagnated till the 1990. Vocaltec’s *Internet Phone* was a breakthrough in VoIP software development and a real motivation for other developers and general public.

Nowadays most telecommunication equipment developers have formed their own vision of new technology deployment. German giant Siemens (SURPASS), Cisco Systems (Architecture for Voice, Video and Integrated Data), Lucent, Telcordia, Nortel, Alcatel, Ericsson, and Iskratel are all actively involved in IP telephony development.

Is the market ready for global IP telephony? Are great expectations justifiable or is all the hype produced with active marketing and Internet revolution? Overview of

Internet growth in the past and assumptions for the future can help us in the search for the answer.

In the year 1950 world community witnessed first trans-atlantic telephone call, in 1960 digital switching became popular, in 1970 we started to program telephone exchanges and software became its expedition to become the most important part of modern telephone exchange. In 1980 Common Channel Signalling System number 7 (SS7) was introduced and made new services, like toll-free numbers, possible. It introduced packet-based transport of Call Control to traditional CSN (Circuit Switched Network). With fast deployment of Internet packet based networks became widely available and it was a dream of many to use them for all their daily voice communication.

Next, importance of IP availability for different VoIP scenarios will be discussed. It will be followed by explanation of packet-based transfer of voice and short example of its usage. In conclusion our views for the close future development topics will be presented.

2 IP availability

For VoIP IP availability is essential. There are many uses of the new technology. If VoIP is not used by end users, IP availability does not present a serious problem. The most frequent use in current telecommunication market is long distance transport of voice streams using packet technology. It enables better exploitation of existing infrastructure, thus considerable cost savings. Most such connections use private IP network which guarantees satisfactory QoS (Quality of Service) provision. Nowadays Internet is not ready to be used for toll quality voice communications.

For the global usage of VoIP — which could be compared to traditional CSN — all regular users would have to have VoIP-enabled terminal. VoIP terminals come in different forms and shapes. They can be specially designed telephones connected to packet network (already available on the market) or properly equipped computers. Most IP telephones are vendor specific. They depend on special software and sometimes even use their own non-standard call control protocols. However, all of them also support standard protocols so they can interact at least with some computer programs. There is a lot of activity in CTI (Computer Telephony Integration) and VoIP in Open Source community, too. In the year 2001, great freely available client and server programs can be

expected.

Statistical data shows (Fig. 1) that number of IP hosts constantly present on the Internet (24 hours a day) will meet the number of telephone subscribers in the year 2005. We can safely say that by then VoIP will be widely available to the end users at their homes and offices. The technology for VoIP is here, we just have to provide better Internet-wide QoS and start using it.

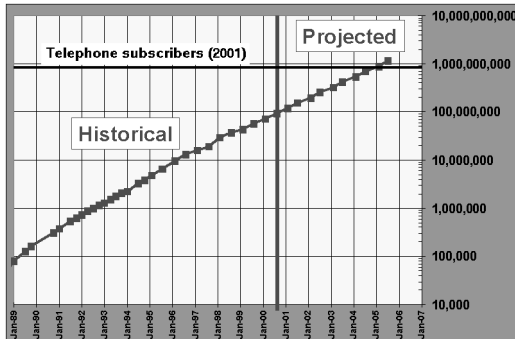


Figure 1: Prediction on the number of Internet hosts in the future [1].

3 Packet based transfer of voice

To understand requirements that have to be met by the backbone and local network architects, we have to look at means of voice transfer over PBN (Packet Based Network). It can be described in five steps (Fig. 2):

1. voice digitalisation,
2. voice processing,
3. voice coding,
4. packet preparation, and
5. sending to the PBN.

First step is the same as in traditional CSN. Analog voice signals generated in telephone handset or microphone connected to computer's sound card are sampled 8000 times per second and coded in 64 kb/s stream. In second step, the voice stream is processed. Algorithms for line echo detection and VAD (Voice Activity Detection) are used. Here first obvious benefit of packet-based voice transfer presents itself. Each telephone conversation includes a lot of silence — moments when one or all of the participants do not speak. With packet-based approach we can save a lot of bandwidth compared to traditional telephony with the omission of “empty” packets. Since users are accustomed to “line-noise” when their peer is not speaking, VoIP terminals can generate background “comfort-noise” to maintain the illusion of a constant stream of noise across the network. Comfort noise can be either generic white-noise hum at a low volume or real noise sampled from the actual call.

Next step, voice coding, is very important part of VoIP communication. There are many factors that influence codec choice. Most prominent are bandwidth availability and required quality. Toll quality can be achieved

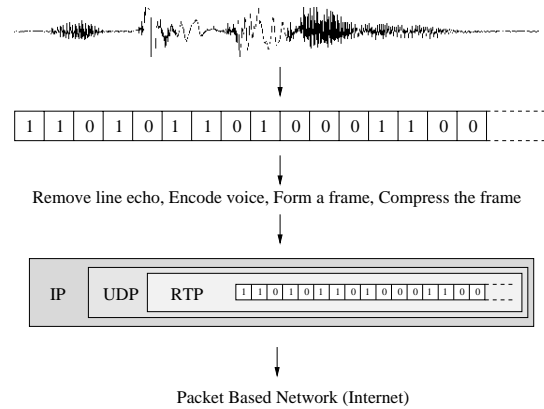


Figure 2: Five steps for voice packet preparation.

with 64 kb/s G.711 which is also used in ISDN (Integrated Services Digital Network). H.323 supports many codecs: G.711, G.722, G.728, G.729A, G.723.1. Support of G.711 is obligatory, all others, even proprietary codecs, can be supported at developers will. This measure should provide at least one common codec between parties using H.323 compliant VoIP terminals.

Table 1: Voice codecs overview.

Codec	Bandwidth requirements	Frame size
G.711	64 kb/s	1 ms
G.722	64 kb/s	1 ms
G.728	16 kb/s	2.5 ms
G.729A	8 kb/s	10 ms
G.723.1	5.3, 6.4 kb/s	30 ms

Table 1 shows that codecs use different frame sizes. H.225.0, which is part of H.323 standard, recommends 20 ms frame sizes. If we decide to use G.711, we would pack 20 frames in one voice packet. Setting the right length for a voice packet is very important.

Packet length dictates ratio between the information part of the packet and an overhead that is introduced by underlying protocols. Let us define “exploit” ratio with $Q_e = \frac{\text{voicedata}}{\text{overhead}}$. If we choose to conform to H.225.0 recommendation and use G.711, IP packet will consist of 40-byte header (IP, UDP, and RTP headers) and 160 bytes of voice data. That demands bandwidth of 80 kb/s and provides $Q_{e.G.711} = \frac{160}{40} = 4$. If we decide to use codec that can operate with lower bandwidth, the ratio may change drastically. With G.728 we form 20 ms with only 8 voice frames (Tab. 1), 5 bytes each, which results in 40-byte header and 40 bytes of voice data. As we can see, although the bandwidth requirement is reduced to 32 kb/s, $Q_{e.G.728}$ has fallen to 1. Introduction of CRTP (Compressed Real-Time Transport Protocol) increased efficiency, with 2 - 4 byte headers. Nevertheless, wise choice of codec and frame length is crucial.

If we assume usage of public IP infrastructure, without QoS provision, packets can take different routes to peers destination. They can arrive in arbitrary order or get lost on their way. For voice packets usage of UDP (User Datagram Protocol) is expected. It is unreli-

able, connection-less protocol for applications that do not want TCP's (Transmission Control Protocol) flow control. When voice packet is lost, the best thing to do, if we do not have any means to restore it, is just to ignore it. Listener will hear a little disturbance in the voice flow, but most users are already accustomed to it from their cell phones. If we would use TCP, lost packet would be retransmitted and the whole conversation would be on hold until successful retransmission. On the other hand TCP is preferred for Call Control. Call initialisation, supplementary services invocation, and billing information is gathered from Call Control, so we want it to be as reliable as possible.

If public network, like Internet, is used for the transport of VoIP packets, packets arrive to their destination at irregular time intervals. This is called jitter and introduces another obstacle in VoIP communication. Currently jitter is solved with the use of special buffers - jitter buffers. They are introducing additional delay in the range of 100 ms. Various researches established that delay of 200 ms is acceptable even for business use, but delays greater than 200 ms showed to be unacceptable. Hitachi Semiconductor (America) Inc. published table (Tab. 2) that presents sources of delay on voice packet path.

Table 2: End-to-end VoIP packet latency and delay

Latency delay source	Typical delay
Recording	10 – 40 ms
Encoding (codec)	5 – 10 ms
Compression (speech coder)	5 – 10 ms
Internet delivery	70 – 120 ms
Jitter buffer	50 – 200 ms
Decompression (speech coder)	5 – 10 ms
Decoding (codec)	5 – 10 ms
Average	150 – 400 ms

That concludes our introduction to basics of VoIP. In the next sections we will present a short example of H.323 and SIP call.

4 H.323 call

The H.323 standard was specified by the ITU-T Study Group 16. First version of the H.323 recommendation was accepted in October 1996. It describes terminals and other entities that provide multimedia communications services over PBN which may not provide guaranteed QoS [2]. The scope of recommendation does not include the network interface, the physical network, or the transport protocol used on the network. It defines functions of the application layer protocols. H.323 entities may provide real-time audio, video and/or data communications. Support for audio is mandatory, while data and video are optional.

H.323 endpoints:

- H.323 terminal,
- Gatekeeper,
- Gateway,

- Multipoint Control Unit (MCU).

The Gatekeeper shall perform alias address to Transport Address translation. This should be done using a translation table which is updated using the Registration messages. Network access is authorised by the Gatekeeper using the H.225.0 messages (ARQ — Admissions Request / ACF — Admissions Confirm, ACR — Admissions Reject). Communication with PSTN devices is accomplished using Gateways. Networks which contain Gateways should contain a Gatekeeper in order to translate incoming E.164 addresses (telephone numbers) into Transport Addresses. The MCU is an endpoint which provides support for multipoint conferences. The network side of the Gateway may be a MCU. A Gatekeeper may also include MCU.

H.323 entities can use Direct or Gatekeeper routed signalling. Later provides means for traffic analysis and billing. Due to space limitations we will show only gatekeeper routed endpoint signalling (Fig. 3).

Before call initialisation, H.323 terminal has to register with the Gatekeeper using H.225.0 RAS (Registration, Admission and Status) messages. Next, Call Setup takes place using call control messages defined in Recommendation H.225.0 (Q.931) (dashed lines in Fig. 3).

Before media streams are established, peer terminals have to accommodate their capabilities with the help of H.245 signals (dotted lines in Fig. 3). Media stream is transported with the use of RTP and RTCP (Real-Time Transport Control Protocol) with one media stream for each direction — full duplex mode.

When one of the peers wishes to release the call, it uses EndSession command to indicate it to the other party. It can be issued by either caller or callee.

5 SIP call

Session Initiation Protocol (SIP) is a signalling protocol for Internet conferencing and telephony. SIP was developed within the IETF MMUSIC (Multipart Multimedia Session Control) working group, with work proceeding in the IETF SIP working group. It is application-layer control signalling protocol for creating, modifying and terminating sessions with one or more participants. These sessions include Internet multimedia conferences, Internet telephone calls and multimedia distribution. It is designed to be independent of the lower-layer transport protocol. It can use TCP or UDP. If compared with H.323 stack, SIP represents Call Control part (H.225.0 - Q.931).

Sessions can be advertised using multicast protocols such as Session Announcement Protocol (SAP), electronic mail, news groups, web pages or directories (Light Weight Directory Access Protocol — LDAP), among others [3].

Unlike H.323, SIP is a text-based protocol such as Hyper Text Transfer Protocol (HTTP). They are very similar in the way they work and transmit information. SIP inherited request-response model and much of its syntax, header fields and semantics.

SIP endpoints:

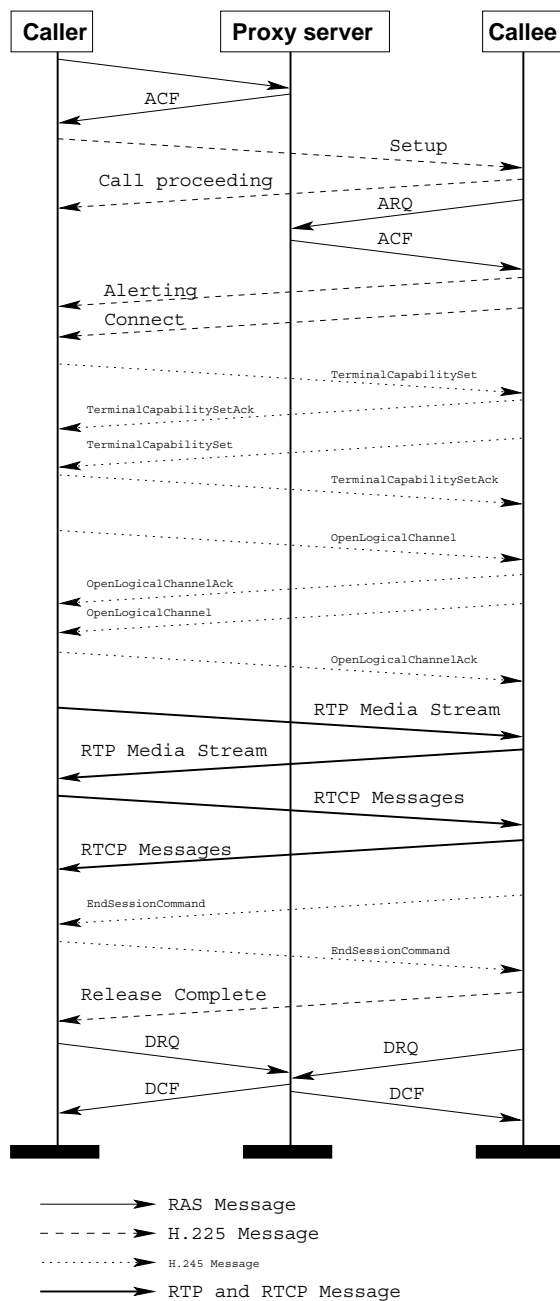


Figure 3: Gatekeeper routed endpoint signalling.

- SIP User Agent,
- Registrar,
- Proxy,
- Redirect.

SIP User Agent (UA) consists of UA server and UA client part. User can register with company's Registrar server. User sip:bv1a@matrix.uni-mb.si can register as sip:bostjan.vlaovic@uni-mb.si at university wide Registrar server, so his SIP-URL can be derived from his e-mail address. When a client wishes to send a request, the client either sends it to a locally configured SIP proxy server or sends it to the IP address and port corresponding to the Request-URI. Request-URI is a SIP-URL as described in [3] or RCF 2396. A callee may move between a number of different servers over time. These locations can be

dynamically registered with the SIP server with the REGISTER request.

A successful SIP invitation consists of two requests, INVITE followed by ACK. The INVITE request asks the callee to join a particular conference or establish a two party conversation. The message body of the request may contain a description of the session using Session Description Protocol (SDP). For the two-party calls, the caller indicates the type of media it is able to receive and possibly the media it is willing to send as well at their parameters such as network destination.

After the callee has agreed to participate in the call, the caller confirms that it has received that response by sending an ACK request. It can be sent through proxy or directly to the callee. A success response must indicate in its message body which media the callee wishes to receive and may indicate the media callee is going to send. The ACK request confirms that the client has received a final response to an INVITE request. Next, media streams are setup with the use of RTP and RTCP in the same way as with H.323.

When the user agent client wishes to release the call, it uses BYE request to indicate it to the server. A BYE request is forwarded like an INVITE request and may be issued by either caller or callee. A party should issue a BYE request before releasing a call (hanging up).

We believe detailed explanation of H.323 and SIP call setup would be necessary. Short introduction was provided with the aim to help users get an practical feeling of VoIP use.

6 Conclusion

VoIP presents one of the most interesting developing technologies. We believe new companies and services will arise in the next couple of years. Introduction of ADSL (Asymmetric Digital Subscriber Line) and other high speed network accesses will provide VoIP to the users homes. Nevertheless IP bandwidth for home use will not be satisfactory for years to come. This is why we believe development emphasis will be on PSTN — IP inter-operation and deployment of new services. Many H.323 implementations will be joined with SIP products and the way we communicate will change for ever.

References

- [1] M. Lottor, Internet Software Consortium, <http://www.isc.org>, December, 2000.
- [2] ITU-T Recommendation H.323, "Packet-based multimedia communications systems," 1998.
- [3] Internet Engineering Task Force, "RFC2543bis SIP: Session initiation protocol," Apr. 2000. (Work in progress).
- [4] B. Vlaovič, Z. Brezočnik, "IP Telephony Today and Tomorrow", *SSGR 2000 Conference*, L'Aquila, July 31 - August 6, Italy, 2000.