Voice Service Interworking for PSTN and IP Networks[1]

M. Hamdi - *Maher.Hamdi@epfl.ch*
O. Verscheure - *Olivier.Verscheure@epfl.ch*
J.-P. Hubaux - *Jean-Pierre.Hubaux@epfl.ch*
Institute for Computer Communications and Applicatins - ICA (EPFL)

Abstract:

In recent years, the Internet has proven its ability to carry real-time data, including voice. Today, a small amount of voice traffic has already been diverted from the Public Switched Telephone Network (PSTN) to the Internet. If it expands, this phenomenon can completely change the rules of the game for telecommunications.

This paper presents an overview of the main technical problems to be addressed for the provision of interoperable services between IP Telephony and the PSTN. The pivotal element of the solution resides in an interworking function. This function is typically implemented in a gateway whose requirements and behavior are here analyzed in terms of signaling and control protocols (control plane), user data transfer (user plane) and management features (management plane).

The presentation is structured around these three planes. The control plane defines the set of signaling protocols to be used in each networking context and the translation between them. Detailed scenarios illustrate the signal translation in the gateway allowing for the establishment of a hybrid phone call. The user plane is responsible for adapting the user data to the properties of each network channel and determines the Quality of Service of the voice call in terms of delay and speech quality. In the management plane, the issues of network, service, security and policy management are discussed.

1 Introduction

IP Telephony is becoming a very successful voice technology as evidenced by the burgeoning market for computer-based telephony products. This was enabled by recent advances in different technologies. In the signal processing field, new speech compression standards allow voice signal to be coded at very low bit rates while keeping their quality acceptable for conversational services. Moreover, the increasing bandwidth in IP access networks associated with the increasing routing capacity in the IP backbone, make it possible to reach an interactivity level similar to that offered by circuit switched networks. In addition, the dramatic growth of IP terminals with expanding processing power, memory and multimedia capabilities, allows IP based voice services to be deployed at a very large scale.

On the other hand, the Public Switched Telephone Network (PSTN) has made very impressive achievements, in terms of coverage, reliability and ease of use. The number of lines is still increasing today, and is about to reach the milestone of one billion. The availability of the service is such that users are used to get dial tone every time they pick up the phone and to be connected to any selected called party. PSTN terminals are also usable by most disabled people and people with limited education. In addition, the telephone network is being extended by cellular networks, which have already attracted more than 200 million subscribers; their growth is almost as dramatic as that of the Internet.

Matching these features with a fully IP-based network is a major engineering challenge. That might last several decades; there is even no consensus today that this will ever happen. Part of the voice services, so far offered by the PSTN, will certainly migrate to an IP based technology. However, we believe that IP Telephony and the PSTN services will coexist for a considerable time.

1. Correspondance to M. Hamdi, EPFL-DI-ICA, Lausanne 1015, Switzerland. Tel: + 41 21 693 5610. Fax: + 41 21 693 6610. Email: Maher.Hamdi@ epfl.ch

For these reasons, the ability to interconnect IP Telephony users to PSTN users is an essential feature. It is the goal of this paper to discuss main interworking aspects between IP Telephony and PSTN voice services.

Two main standardization approaches are being carried out for IP/PSTN interworking. In the IP world driven by the IETF (Internet Engineering Task Force), interworking with the PSTN has been the result of a logical extension to the IP telephony service, which is seen as one of many IP applications. AVT, IPTEL, MMUSIC and PINT are the main IETF working groups concerned with IP Telephony. In the telecommunications world, the International Telecommunication Union - Telecommunication Standardization Sector (ITU-T) and the European Telecommunications Standards Institute (ETSI) are the main contributors in terms of standards and pre-standard documents. The ITU-T has initiated various standardization activities (e.g., [11] [12] [13] [14] [15]) that captured the attention of most of the industrials involved in the field. Related to these standards, the ETSI project TIPHON (Telecommunications and Internet Protocol Harmonization Over Networks) undertook the effort to identify additional technical agreements required for the interoperability between IP networks and circuit switched networks [4]. Some industrial consortia such as the International Multimedia Teleconferencing Consortium - IMTC (through its VoIP- Voice over IP- group) also provide recommendations related to the implementations interoperability that is required in a multi-vendor context [8].

In this paper, we analyze the main requirements for interworking between IP Telephony and the PSTN services. Illustrations are based on the H.323 standard. For clarity purpose, the interworking features are organized in three planes: the Control Plane, the User Plane and the Management Plane. The Control Plane interworking defines the set of signaling protocols to be used in each networking context and the translation between them. The User Plane interworking is responsible for adapting the voice data to the properties of each network channel and determines the Quality of Service of the voice call in terms of delay and speech quality. In the Management Plane, we present a brief overview of main management aspects in the context of hybrid voice services.

The paper is organized as follows: Section 2 defines hybrid voice services and gives basic communication scenarios for IP/PSTN interworking. The PSTN/ISDN protocols and H.323 systems are briefly reviewed. The interworking features in the Control Plane are described in Section 3, where we discuss signaling adaptation, addressing and media control functions. User Plane interworking is discussed in Section 4. The impact of end-systems and network design is analyzed in terms of speech quality and communication interactivity. In Section 5, we discuss some aspects of the Management Plane and related open issues. Section 6 concludes the paper.

2 Voice Service Interworking

Interworking of IP and PSTN voice services can be considered as a part of a much bigger effort undertaken by standardization bodies in the field of network and service interworking [3] [23] [5]. The most obvious interworking scenario between IP and the PSTN is when the PSTN connection is used as a lower data layer by the access part of an IP network (e.g. dial-up access to an Internet Service Provider). We rather focus on service interworking, and more specifically, on interworking of voice services. In the context of PSTN and IP Telephony services, interworking is the ability to offer a broader service that results from their peer juxtaposition. For the remainder of the paper, voice services resulting from this interworking will be referred to as *hybrid voice services*. More concretely, hybrid voice services provide connectivity between users of both networks as well as between users of the same network given that part of the communication uses the service of the other network. Therefore, hybrid voice communications involve both PSTN and IP voice services and/or both types of terminals.

In this section, we describe five basic scenarios for voice communications. We consider voice services over the PSTN, the IP network as well as hybrid combinations. Also, we give a brief overview of the H.323 standard.

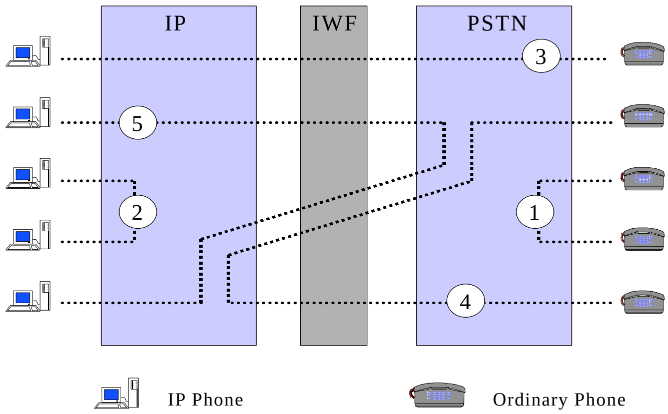


Figure 1: Voice Communication Scenarios

2.1 Five scenarios for voice communications

Figure 1 illustrates five basic voice communication scenarios. Hybrid voice services are represented by scenarios 3, 4 and 5. In these scenarios, an Interworking Function (IWF) is needed to perform all protocol conversions and data adaptations. The IP and PSTN areas represent a protocol concept and do not necessarily involve a real network. Therefore, an IWF device may be used to connect two networks (i.e., a network adaptor) or a terminal to a network (i.e., a terminal adaptor).

For voice services, the IWF provides the following mechanisms:

- *Signaling Adaptation:* it consists in the processing and translation of incoming signaling messages. It mainly concerns the call setup and clearing phases.
- *Media Control:* it consists in identifying, processing and translating service specific control events that may be generated by the user or the terminal.
- *Media Adaptation:* it consists in adapting the voice data to the data transfer channel of the downstream network.

2.2 The PSTN Voice Service

In scenario 1, two standard phone sets are connected via the PSTN. Although well known by the communications community, we briefly review the main PSTN characteristics that will be crucial to further discuss interworking concepts.

The PSTN core network is based on a circuit switched network where each circuit corresponds to a 64 kbits/s digital channel. A PSTN terminal can either be digital or analog. Standard phone sets are attached to the PSTN by means of an analog access network, which merely corresponds to the set of subscriber loops (the copper wires that link the customers to the Central Office). On an analog access network, voice is transmitted as a 3 kHz wide-band analog signal and gets digitized at the access switch. In this case, signaling capabilities on the analog part of the access network (e.g. address notification) are reduced to in-band coding of DTMF tones (DTMF stands for Dual Tone Multifrequency).

The ISDN allows voice terminals to have digital access to the PSTN. In this case, a digital voice terminal (or an analog terminal attached to an adaptor) initiates a signaling dialog using Q.931 [18] (or the Digital Subscriber Signaling System N. 1, DSS-1) to connect to the network via a 64 kbits/s digital channel. Signaling inside the digital core network is based on the Signaling System N. 7 (SS7) [22]. An ISDN terminal seamlessly calls an analog PSTN terminal and vice-versa. A unified addressing system is defined in ITU-T Recommendation E.164 [9].

Finally, one essential feature of the PSTN is its service creation and control capabilities referred to as Intelligent Network (IN) [17]. Basic services such as call forwarding rely on the IN architecture.

2.3 Voice Services over IP

Scenario 2 illustrates what is generally referred to as IP telephony. IP telephony follows the IP paradigm: all service-specific processing and protocols, such as signaling and media coding, are pushed to the end-systems and are transparent to the network. Applications may be built on top of TCP or UDP depending on whether they are loss-sensitive or time-sensitive respectively. For example, the TCP transport protocol is used to carry the signaling stream since the signaling channel has to be error-free. However, because of its intrinsic timing constraint, voice data is usually transmitted over UDP. The time-continuous property of voice signals requires that the transport channel ensure the appropriate streaming needed for data resynchronizing at the receiver. For this reason, the Real-Time Protocol (RTP) [26] is used. The sequence numbering field of RTP packet headers is used to re-order the receiving packets in case of out-of-sequence delivery (UDP does not ensure packet sequencing), the time-stamp field indicates the temporal play-back position of the data payload. In addition, RTP allows the receiver to identify the media coding type (i.e., which voice coding standard has been used at the coder side).

As far as end-users are concerned, personal computers (PC) are the most common IP terminals. The processing and control parts of an IP Telephony terminal are therefore usually implemented in software. However, a standard telephone set can also be connected to an IP Telephony service by means of a network adaptor that provides a minimal set of the required protocols. This has the advantage to potentially reach a much larger number of users than PC holders.

The ITU-T Recommendation H.323 [11] and its related set of standards for packet based multimedia communications [12] [13] [14] [15] - in addition to the several related efforts carried out by the ETSI, IETF and the IMTC - certainly constitute the most advanced framework that covers essential IP Telephony issues. Although it is not our goal to present a tutorial on H.323, a brief description of the standard is required for the following discussion on interworking. The presentation is restricted to the basic voice aspects of H.323; data and video communications as well as multi-point aspects are not covered.

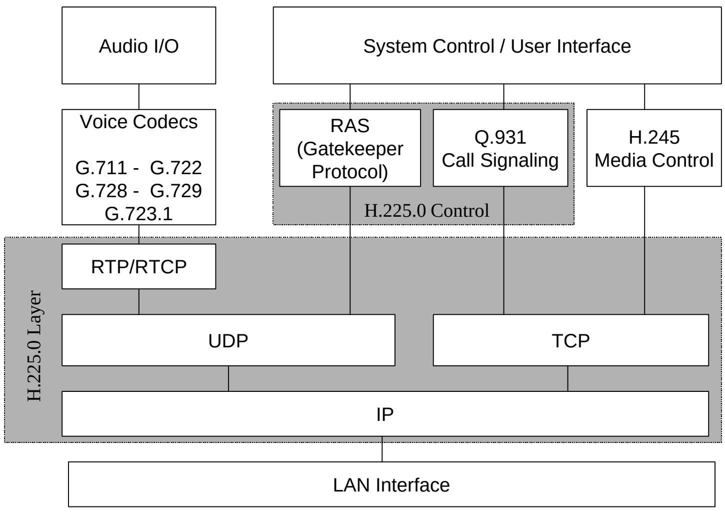


Figure 2: H.323 Voice Terminal

H.323 Systems

The H.323 standard defines three types of equipment: *gatekeepers*, *gateways* and *terminals*. The *gatekeeper* is an optional equipment that provides call control services to the terminals. Examples of such services are address translation, admission control, call authorization and directory services. The RAS (Registration, Admission and Status) protocol defined in H.225.0 is used to communicate between a terminal and a gatekeeper.

The *gateway* is responsible for providing all translations necessary for transmission formats and control procedures between the IP supported portion and the PSTN/ISDN part of hybrid calls. As gateway functions are more related to hybrid calls than pure IP calls, they will be discussed in the Section 3.

The *H.323 terminal* components are described in Figure 2. A terminal may support several standards for voice coding. The G.711 codec (used in ISDN) is however mandatory for all terminals. The H.225.0 Recommendation specifies the use of logical channels based on the RTP/UDP/IP protocol stack to transfer coded voice data. The *system control part* of a terminal is composed of three protocols:

• The RAS signaling function is used for the dialog between a terminal and a gatekeeper. The associated channel, called RAS channel, uses the UDP/IP protocol stack. A main function of the RAS channel is to allow the terminal to be attached to a gatekeeper by registering itself. Registration basically results in an update of the gatekeeper's address translation table. This allows other terminals to locate the registered terminal and to determine its transport address in order to initiate a call-signaling channel.
• The call signaling between two H.323 terminals is based on Q.931 messages. The call-signaling channel uses a TCP/IP protocol stack. The call-setup phase consists in sending a Setup message to the destination. The setup phase is considered successful upon reception of the Connect message from the called terminal. The next phase is the establishment of the H.245 channel.
• The H.245 protocol defines end-to-end control messages used for capability negotiation (e.g., the supported codecs), opening and closing of logical channels, flow control messages, etc. The H.245 control channel is a reliable channel based on TCP.

Figure 3 shows an example of a control protocol diagram between two H.323 terminals. A description of some of these messages is given in Section 3.3. Finally, it should be noted that H.323 defines a Fast Connect method in order to alleviate the initiation phase in basic and simple calls. The H.245 dialog is then replaced by additional information elements in the Q.931 messages so that, upon reception of the Connect message, all needed voice channels are activated.

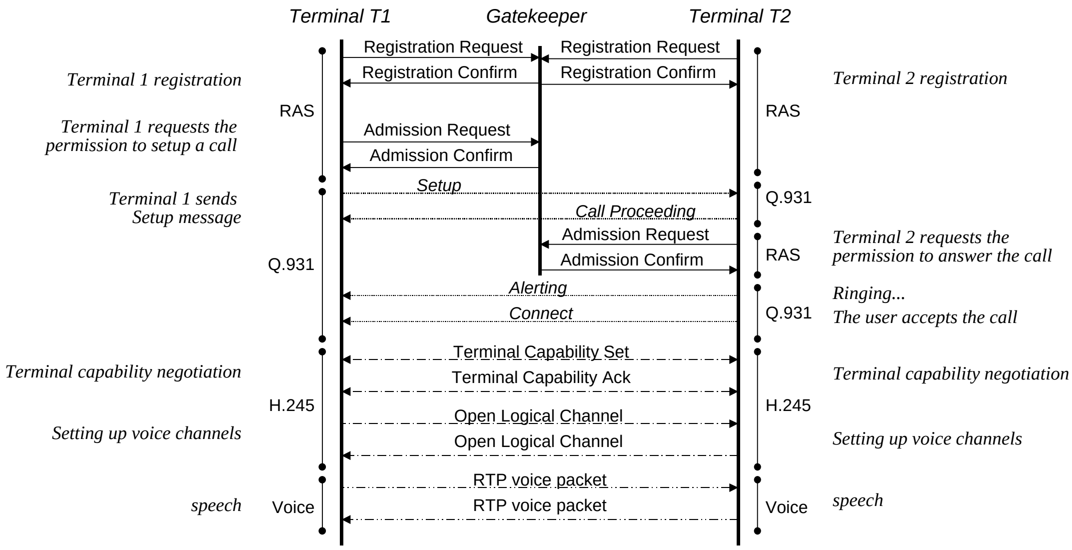


Figure 3: Diagram of H.323 Control Protocols

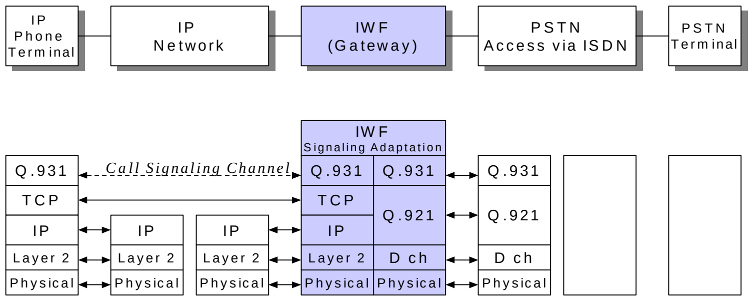
2.4 Hybrid Voice Services

In scenario 3, the two terminals involved in the call use different protocol stacks to communicate with their access networks. The protocol conversions occur at the networks' boundary. Two terminals, of different type in this case, communicate with each other to ensure an ad-hoc voice service to the end-users. Scenario 3 requires both the mapping of media and media control channels and the mapping between signaling protocols.

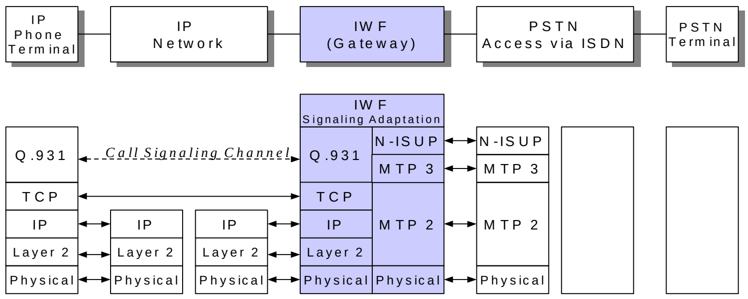
In scenarios 4 and 5, the same protocols are used at the interface of each terminal, but a different protocol is used between them. The protocol conversions (at least twice) in both directions take place at the boundaries of the traversed networks and the presence of another network in the middle should be transparent to end-users. In these scenarios, both the mapping of media and media control channels and the mapping between signaling protocols are generally required. However, mapping between signaling protocols can be avoided in some configurations. In particular, when the IP network is only used as a backbone network (scenario 4), all PSTN/ISDN signaling information can be transferred transparently through the IP network.

The gateway is the equipment that generally hosts the interworking functions. However, in the H.323 standard, the gatekeeper may also be involved in some interworking functions such as address translation. In the next sections, we will generically call "a gateway" the equipment in charge of all interworking functions.

In addition, we will only consider the case where the PSTN/IP gateway is connected to the PSTN via an ISDN access. Interworking issues between ISDN networks and analog PSTN terminals are not covered in this paper.



(a) - The IWF appears as an ISDN Terminal to the PSTN



(b) - The IWF appears as a Network Node to the ISDN.

Figure 4: The IWF in the Control Plane

3 Signaling and Control

In this section, we show how call connections are set up and how control commands are conveyed during a communication.

3.1 Signaling Adaptation Functions

If two different signaling protocols are used in the interconnected networks, then the IWF should translate the signaling messages in such a way that the end-to-end call can be completed. In the H.323 gateway, Q.931 is used in both the IP network and in the ISDN access. However, the Q.931 signaling channel between an IP terminal and the gateway is terminated at the gateway (i.e., Q.931 messages are processed in the gateway and not simply forwarded). A peer Q.931 channel is then used to support the call control on the PSTN side. This is mainly due to the fact that H.323 has defined a particular use of Q.931 messages, so that there is not necessarily a perfect correspondence with the ISDN use of Q.931 [12] [13]. Figure 4-a shows the IWF protocol stacks in the Control-Plane in the case of Scenario 3.

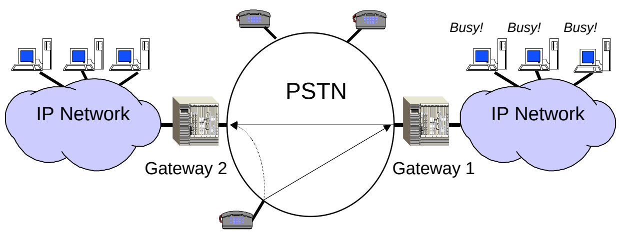


Figure 5: SS7 Interoperability: A Call Diversion Scenario

3.2 SS7 Interoperability

For historical reasons, IP/PSTN gateways are usually seen as administrative boundaries between a network provider (usually the operator) and a network customer (usually a company or an Internet Service Provider). For this reason, they are connected to the network as *terminals*. However, the gateway can be connected as a *network node* to the PSTN, to have access to its SS7. Consider, for example, the scenario depicted in Figure 5. Two IP-Telephony based call centers are shown, each of them is connected to the PSTN through gateways. The two call centers are combined to form a single virtual, distributed call center; if all the agents in one call center are busy, the calls are to be diverted to the other one. If the gateways do not have access to the SS7 network of the PSTN, then such a call diversion requires terminating the call at the first gateway, and re-initiating a call from the first gateway to the second one. This would tie up two PSTN ports of the first gateway, use up two voice circuits in the PSTN, and potentially introduce a high delay due to the convoluted route that the voice signal follows. On the other hand, if the first gateway has access to SS7, then it can simply divert the call to be directly terminated at the second gateway, thereby avoiding the above inefficiencies. In this way, the two call centers can seamlessly be joined to form a virtual call center, which can be called at a common phone number. In this case, the gateway needs to implement the N-ISUP (Narrowband ISDN User Part) protocol. Figure 4-b shows the protocol stack needed for scenario 3 where with the gateway connected to the PSTN as an ISDN node.

SS7 is central to the operation of the PSTN. Therefore, the telecommunications companies are very reluctant to expose their SS7 network to gateway owners. A more acceptable approach is to provide an SS7 access to a *signaling gateway*, which would control one or more *media gateways*. The signaling gateway would then reside on the premises of the telecommunications company, and would communicate with the media gateways via a specific protocol. Work is in progress in IETF and ITU-T Study Group 16 in order to standardize such a protocol. Several candidate protocols exist, such as the Simple Gateway Control Protocol (SGCP), the IP Device Control (IPDC) protocol suite, and the Reliable Gateway Control Protocol (RGCP). Another proposal in consideration is to use the H.323 signaling protocols for this purpose.

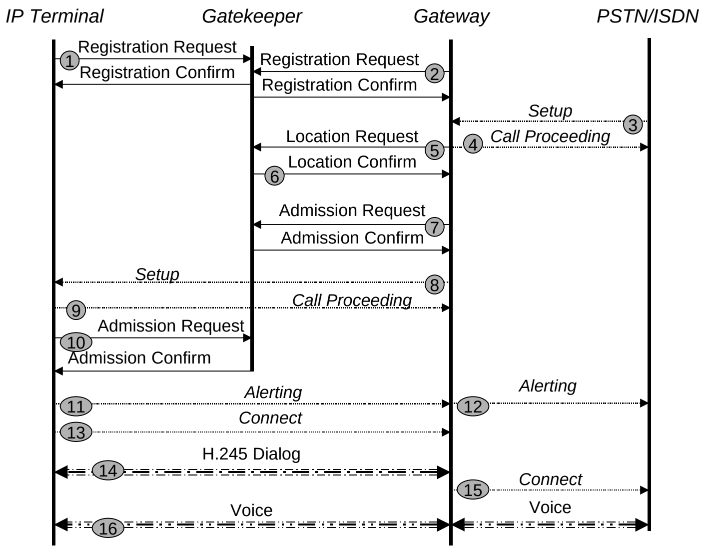


Figure 6: Example of Call Control in a Hybrid Voice Communication

3.3 Addressing

In the IP world, terminal addressing is generally based on alphanumerical streams whose resolution and directoring are based on hierachically organized servers [21]. Similar addressing schemes for IP telephony are provided by the Session Initiation Protocol (SIP) [7] defined by the IETF. However, as a requirement of service interworking between the PSTN and IP, each PSTN user should be able to call an IP attached user and vice-versa. When the call is initiated from an IP terminal towards the PSTN, the E.164 destination address can be easily sent to the gateway and then across the PSTN. Such is the case in both H.323 and SIP.

The problem is more complex when the caller is a PSTN terminal and the destination is an IP terminal. This is partly due to the limited dialing capabilities of standard telephone sets, particularly if only an alphanumerical type of addressing is defined for the destination.

One of the crucial questions is whether the numerical expression of an IP address can be explicitly used in the identification of the IP terminal. An important requirement for service interworking is that the calling user should be oblivious of the network (PSTN or IP network) to which the callee is attached. The ITU-T approach to solve this problem is to allow a H.323 terminal to be identified by several address aliases of different kinds, typically, an E.164 address and an email-like address [11]. Such an approach generally requires specific address translation, resolution and registration services, which in H.323 are typically performed by the gatekeeper.

Figure 6 shows an example of call control scenario with address resolution. A PSTN terminal initiates a call to an IP terminal using the E.164 address alias. The steps of this scenario are the following:
1- The IP terminal registers with the gatekeeper by giving a network address, aliases of the network address and the transport address of its signaling channel (i.e., the TCP port number and IP address). Examples of network address aliases are *user@host* and an E.164 address. The terminal sends as many Registration Request messages as necessary to register all its address aliases.
2- The gateway registers with the gatekeeper in the same way.
3- The gateway receives a Setup message from the ISDN access switch. This message contains the E.164 address of the calling PSTN/ISDN terminal and the E.164 address of the called IP terminal.
4- The gateway sends back the Call Proceeding message to indicate that the call is being processed.
5- The gateway sends a Location Request message to the gatekeeper asking for the channel signaling transport address of the called terminal; the E.164 address of the called party is provided in the message.
6- The gatekeeper sends back a Location Confirm message containing the required transport address.
7- The gateway asks for permission to set-up the call by sending an Admission Request to the gatekeeper. Upon reception of the Admission Confirm message, the gateway is ready to start the Q.931 set-up phase.
8- The gateway sends a Setup message on the signaling channel of the destination IP terminal.
9- If the terminal is alive, a Call Proceeding message is sent back.
10- The terminal asks for permission to set-up the communication.
11- The terminal sends an Alert message to the gateway indicating that the called user is being alerted of the incoming call. This may correspond to the usual ringing signal.
12- The Alert message is forwarded to the ISDN part.
13- The terminal sends a Connect message to the gateway indicating that the call is accepted. The Connect message contains the transport address needed for the establishment of the H.245 channel.
14- The terminal and the gateway initiate the H.245 dialog for capability exchange and logical channels establishment.
15- After the media channels are activated between the terminal and the gateway, the latter sends a Connect message to the ISDN calling party indicating that the voice communication can start.

It should be noted that this diagram depicts a typical scenario, but there exist shorter scenarios that use the Fast Connect procedure [11].

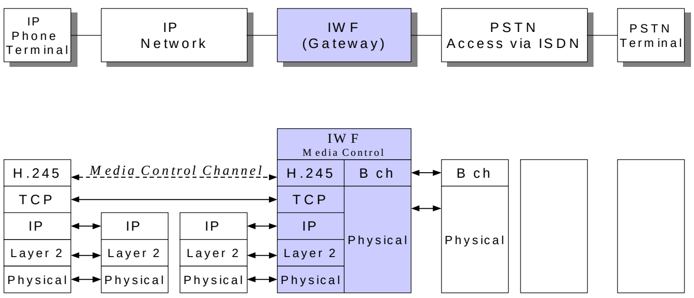


Figure 7 The IWF in the Control Plane: Carrying DTMF signals

3.4 Media Control Functions

Once the connection is set up, the media control channel is used to carry all control information generated by the user or by the terminal. For voice communications, the main user-level control informations are the DTMF tones used, for example, to interact with a voice server. Carrying these signals over a hybrid connection requires a particular attention. The standard compression techniques used today for low bit-rates introduce enough distortion to corrupt DTMF analog tones making the receiving end-system unable to correctly decode the original signals. Therefore, they need to be separated from the audio signal at the sender (if it uses a classical terminal attached to an adaptor) or at the gateway, and conveyed separately to the receiver.

Two approaches have been recommended by the VoIP Forum for carrying DTMF informations. The first is to carry them in-band via RTP using a dedicated payload format. This has the advantage that the tones remain temporally synchronized with the speech. However, packet delivery is not guaranteed because of the unreliable transport protocol UDP. Although packet loss can be kept very low in well-engineered networks and has negligible impact on voice quality, the loss of a DTMF tone can result in severe service malfunctions at the user level. The second solution uses out-of-band transport of DTMF signals on a separate and reliable media control channel. The drawback of this approach is that the signals loose their exact temporal position in the voice stream. This latter approach has been recommended for H.323 systems, i.e., using the H.245 channel. In this case, the protocol stack for Media Control Interworking Function is given in Figure 7.

4 Media Adaptation Functions

The major User Plane issue is to maintain the Quality of Service (QoS) required for voice connections. Instead of worrying about the quality of the transmitted bits, we focus on the quality of information delivered to the end-user. Two main factors may influence the QoS experienced by the end-user:

- The *end-to-end speech quality,* which may be affected by both the successive encoding/transcoding operations and the packet loss due to network congestion.
- The *end-to-end delay,* which mainly impacts the interaction between the participants of a conversation. It results from the coding/decoding process, packetization and queuing delays.

On the IP network side, the service provider tries to accommodate a maximum number of voice connections at a time. Therefore, a key question arises: what are the appropriate mechanisms to be employed within both the end-systems (including the gateways) and the network in order to optimize network utilization while maintaining the desired QoS for the end-users?

In this section, we first give a brief description of the techniques that are being implemented in IP networks in order to provide a certain service guarantee. Then we analyze how the end-systems may influence the user-oriented QoS. We focus on the trade-off among bandwidth, delay and computational complexity.

4.1 QoS and IP-based Networks

While the PSTN network insures a fixed delay and no-loss guaranteed service, this is not necessarily the case for IP-based networks. Indeed, services currently experienced on the Internet are *best-effort* services. They are characterized by the absence of any QoS specification at all. However, IP telephony applications will definitely need some kind of quality guarantees in terms of absolute delay, delay jitter and packet loss.

The Integrated Services (*IntServ*) architecture was designed to provide a set of extensions to the best-effort traffic delivery model. For this purpose, it defines two classes: the Guaranteed Service and the Controlled-Load Service [1].

- Guaranteed service (GS) provides a lossless transfer with tight delay bounds for flows that conform to the parameters negotiated at the connection setup.

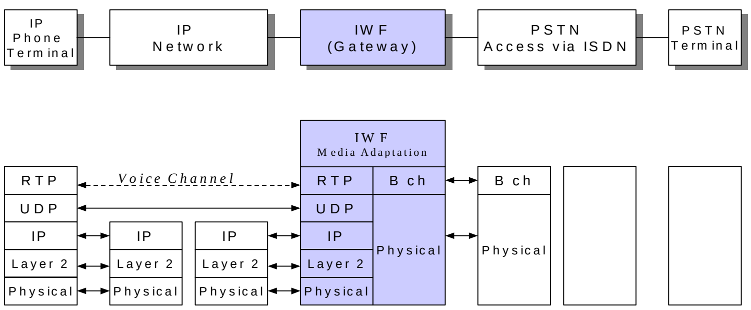


Figure 8: The IWF in the User Plane

• Controlled load service (CLS) yields a quality corresponding to a lightly loaded IP network at best-effort; it is not expressed quantitatively. The admission control is based on the peak rate declared by a session initiator and on measurements of the load in the network. This could lead to a higher network efficiency compared to admission control based only on declared source descriptors.

Both GS and CLS connections can be established by the Resource ReSerVation Protocol (RSVP) signaling [27]. However, RSVP has some weaknesses that considerably undermine its wide deployment, mainly due to the soft-state reservations paradigm and the exponential growth of the reservation state tables.

In order to get around the weaknesses of the solutions proposed by the IntServ group, a new group, called Differentiated Services (*DiffServ*) group, was formed. They suggested that instead of maintaining the state of each and every flow, why not discriminate the packets according to their precedence? The precedence of a packet is indicated by the three first bits of the IP Type-of-Service field. This idea led to the concept of differentiated services, which also has the advantage of being "easily" implementable in existing networks.

As previously mentioned, the objective of a service provider is to increase the network efficiency (reduce service cost) by accommodating as many voice connections as possible. This leads to higher packet loss ratios and delays. The sensitivity of IP voice services to data loss strongly depends on the mechanisms implemented in the end-systems.

4.2 QoS and End-Systems

The heterogeneity of networks causes voice traffic to be handled differently. Indeed, in the PSTN, voice connections generally operate at the standard rate of 64 kbits/s (pulse code modulated, PCM, signal or G.711). However, there is no need to keep such a high bandwidth connection within the IP network. Rates ranging from 5.3 kbits/s (i.e. G.723.1) to 8 kbits/s (i.e. G.729) will usually be more appropriate. The transcoding (PSTN to IP network) process occurs in the gateways. However, a lower bit rate will generally involve a lower signal quality and higher delays. Indeed, while both the G.729 and G.711 coding standards provide a voice quality comparable to the usual telephone service quality (*toll quality*), an encoder based on the G.723.1 standard outputs a quality lower than toll quality. The introduced delay results from both a higher *processing delay* and an increasing *packetization delay*. The processing delay is the delay required to run the encoding algorithm on the uncompressed voice signal and create a stream of bytes ready to be sent to the packetization layer. The packetization delay represents the time needed to form a packet of compressed voice information of a given size. Therefore, when decreasing the encoding bit rate, the service provider can accommodate more voice connections at the expense of increasing signal distortion and delay.

As stated in Section 2, the RTP/UDP/IP protocol stack is used for the delivery of delay- and loss-sensitive services over packet networks. In such a scenario, every single packet contains 40 bytes of pure header information (assuming no header compression technique is used). There is thus an inherent trade-off between packetization delay and payload-to-header ratio (channel utilization): the higher the payload-to-header ratio, the higher the packetization delay for a given encoding standard. For example, if the G.723.1 standard is

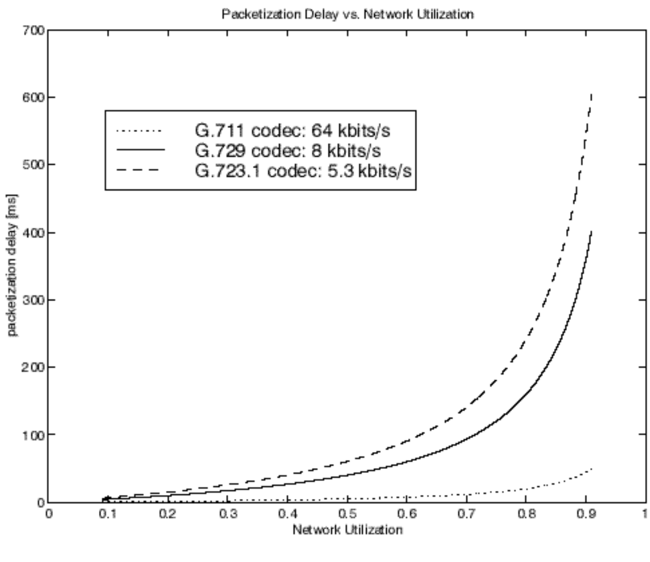


Figure 9: Packetization Delay vs. Network Utilization

used, 60 ms are necessary to collect 40 bytes of voice information (this corresponds to a 50% channel utilization). Figure 9 illustrates the evolution of the packetization delay versus network utilization. However, the packetization delay can be dramatically reduced by multiplexing several voice connections in the same IP packet. A recent Internet draft [25] proposes to perform this multiplexing at the RTP layer (e.g. gateway-to-gateway communication in scenario 4).

The combination of processing, packetization and queuing delays forms the *end-to-end delay* perceived by the end-user. An increasing end-to-end delay may lead to a better service implementation from the service provider viewpoint. However, this end-to-end delay, if strictly lower than 400 ms, should not affect the interaction between the participants of a conversation. Delays up to 150 ms require echo control but do not compromise the effective interaction between the users.

Equivalently, the distortion introduced by both the successive encoding/transcoding processes and the data loss due to network congestion affect the *end-to-end speech quality*. This quality must be equal or close to the toll quality. Mechanisms such as error correction and error masking techniques should be used in order to tolerate higher data loss while providing the same service quality. For example, G.723.1 interpolates a lost portion of the voice signal by simulating the vocal characteristics of the previous portion and slowly damping the signal [20]. The efficiency of an error masking scheme decreases when the number of packets lost at a time increases. Also, Forward Error Correction (FEC) schemes have been proposed to alleviate loss bursts of a small number of packets. An RTP payload type for streams with FEC is being defined by the IETF [24]. It should be noted that FEC introduces some predictable delay.

Although the relationship between all the factors influencing the service quality and network efficiency is intrinsically complex, it is the key to implement an optimal voice service over IP networks.

5 Management Issues

The Management Plane generally covers different layers such as, device, system and service management. A complete management framework for hybrid services is not yet defined because of the different IETF and ITU-T management approaches. Instead, we discuss hereafter, some important management features and related open issues.

5.1 Service Creation

The ITU-T has defined Recommendation H.450.1 to specify how new services (called supplementary services) can be added to H.323 systems. Two supplementary services are already defined in H.450.2 (Call Transfer) and in H.450.3 (Call Diversion) and few others are still under study. The main advantage of standardizing the supplementary services is to ensure their interoperability across different service providers.

IP Telephony clearly offers a much more flexible and open environment for service creation because it relies more on software based and intelligent end-systems than on network nodes. An important issue in voice service management is the development of powerful service creation environments as well as protocols and APIs for uploading these new services in the terminals. Presumably, an approach similar to the Intelligent Network, but more open and flexible is needed for hybrid service management. A Java-based approach has already been proposed for this purpose [6].

5.2 OAM-like features

Up to now, most of the effort spent to achieve IP and PSTN service interworking was focused on the Control and User Planes. The Management Plane interworking is still an open issue. It is however a determinant aspect for the long-term viability of hybrid voice services, especially for operators who are introducing the IP technology in their core or access network. This is because global and unified management operations (e.g. performance monitoring, failure detection) are necessary to ensure seamless service quality for end-users. For the ITU-T standardized networks (ISDN, B-ISDN, Frame Relay), these management aspects are referred to as Operations Administration and Monitoring (OAM). Examples of OAM features are detection of failures and defects, loopback activation and performance measurements. It is then essential that such procedures could be defined and standardized for IP Telephony as well. In particular, for the H.323 standard, some tools that can perform OAM-like features are the following:

- The exchange of the RAS messages Information Request (IRQ) and Information Request Response (IRR) between the gatekeeper and terminals provides a way to detect failures in H.323 equipment.

- The H.245 message Round Trip Delay Request allows an end-point to measure the communication delay in real-time. This information can be used for various management decisions.
- The H.245 message Maintenance Loop Request allows an end-point to setup a connection, in the loopback mode, with another end-point. This procedure can be used for remotely testing the connection availability.

However, more work remains to be done to provide a unified OAM approach for managing a hybrid IP/PSTN platform.

5.3 Network and System Management

The philosophy of network and service management is very different in the PSTN and in the Internet. The telecommunications community has defined and to a large extent implemented an architecture called the Telecommunications Management Network (TMN) [16]. It is a heavy-weight solution, in which the various network components are represented in a sophisticated object-oriented model. In this architecture, five functional areas are identified: Fault, Configuration, Alarms, Performance and Security.

For IP Telephony systems, the ITU-T study group 16 is working on the definition of several MIBs (Management Information Base) for the various H.323 components (e.g. see [19] for the gatekeeper's MIB). It is absolutely not clear how the TMN functional areas will interface with the SNMP-based management of the Internet for the successful provision of hybrid voice services.

5.4 Policy Management

From a service management viewpoint, the SNMP-based approach for device-level management, has some limitations. For example, it is difficult for a network manager to map the desired network behavior into individual device configuration parameters, especially considering that the desired behavior may depend upon many dynamic factors such as the traffic conditions, the time of the day, the date, and the network topology. Furthermore, device-level management allows only very limited possibilities for altering the network behavior based on the sender or the receiver of the information.

In order to overcome these limitations, the networking community has been working on *policy management*, which can be considered as a level of abstraction above device management. Policy management allows the network managers to specify device-independent policies that describe the network behavior in terms of security, quality of service, accounting/charging, etc... Such policies are stored in policy servers, and downloaded to individual devices as required. Therefore, a protocol is required to communicate between the network devices and the policy servers. The IETF defined COPS (Common Open Policy Service) protocol can be used for such purposes[1].

By using the RAS protocol, the H.323 devices request from the gatekeeper various services such as zone registration, call admission control, address translation, etc... The current gatekeeper implementations make these decisions locally, without taking into consideration the network conditions. However, a policy-aware gatekeeper would consult the policy server for making its decisions. This provides a great flexibility for policy implementation, which is no more limited by the gatekeeper capabilities. Therefore, the gatekeeper would be the natural front end to the policy server for the H.323 systems.

The policy management concept is still in its infancy, and many of its components are yet to be developed and standardized. The efforts for standardization are being carried out by bodies such as the Desktop Management Task Force (DMTF) and the IETF [2].

5.5 Security

Security in voice communications is gaining more and more interest for both the PSTN and IP Telephony. User/terminal authentication as well as communication privacy are the most frequently required security features. A number of devices are commercialized today to secure telephone (and fax) communications on the PSTN. On the side of IP networks, many proposals have been implemented at the network, transport and application levels. Concerning H.323 systems, a security framework has been defined in Recommendation H.235. However, the security protocols and algorithms have not yet been standardized.

1. While COPS has originally been defined for RSVP policies, it is applicable to other types of policies as well.

Usually, security mechanisms are negotiated between the two end-points of the communication. However, in hybrid voice services, voice channels as well as signaling channels are also terminated at the gateway. How will this “third-party” be involved in the authentication procedure is still an open question. This rises notably the issue of key management in a hybrid environment, a problem that seems to have been overlooked so far by the research community.

5.6 Dependence on the regulatory framework

For historic and strategic reasons, the telecommunications regulatory bodies have devoted a lot of attention to the voice service. Even if today voice services are being liberalized in most countries, the way they are billed is still under strong political monitoring. Therefore, how profitable the business of IP Telephony can be, will also depend on what the legislation is going to decide in the years to come.

In spite of these open questions, the migration of voice services from the PSTN to IP networks is absolutely crucial. Even if data bits are becoming more numerous than voice bits, voice still represents more than 70% of the revenue of the whole communication business. A wealth of advanced multimedia services can be envisaged on this basis.

6 Concluding Remarks

In this paper, we have given an overview of the different technical issues involved in the provision of voice services over hybrid PSTN and IP networks. Due to the various aspects mentioned below, it is quite difficult to predict the pace at which this new technology will be accepted. There are a number of problems that still need to be solved; we discuss them briefly.

Complexity

The provision of voice services over the conventional PSTN is already extremely complex, and has led to highly sophisticated switches running programs of millions of source code lines. Five communication scenarios have been discussed in this paper. This means that all of a sudden, the combinatorial complexity of voice services could be increased by a factor of five. It would be too optimistic to believe that this problem is solved by the intelligence-at-the-edge paradigm of the Internet. Indeed, as we have seen, many functions have to be implemented in the gateways and gatekeepers, which are centralized devices of the network. This means that even if the communications are established (and billed) properly for the five basic scenarios, it does not prove that this will be the case in more complex configurations; for example, what happens if a user wants to establish a call from a conventional phone to an IP terminal, which happens to have its calls forwarded to a cellular phone? If this works, will this call be properly charged?

Quality of service

As we have seen, in a hybrid call the user data have to go through a number of transcoding operations. It has not been proved that the user-perceived quality of service will be acceptable in a widely deployed hybrid service. Moreover, if at least one of the terminals happens to be mobile, the combination of the wireless problems with packetization delays and degradations due to transcoding can be quite a challenging problem.

Ease of use

It is clear that users appreciate the ease of use of universal communication systems such as the telephone and the electronic mail notably because of the simplicity of the addressing principle. However, in the case of hybrid voice services, this simplicity will disappear; this is due mainly to the fact that end-users will be in some way aware of the existence of intermediate devices (gateway, gatekeeper). As we have seen, there are proposals to make users oblivious of these issues by mechanisms based on aliases; however, these mechanisms can be highly vulnerable e.g. to changes of telephone or Internet service providers.

Acknowledgement

The authors would like to thank Ismail Dalgic and Constant Gbaguidi for their contribution to this work.

References

[1] P. Almquist. Type of Service in the Internet Protocol. IETF RFC1349. July, 1992.

[2] J. Boyle et al.. Common Open Policy Service Protocol. Internet Draft. August 1998.

[3] S. Dixit and S. Elby. Frame Relay and ATM Interworking. IEEE Communication Magazine. Vol. 34, N.6, June 1996.

[4] ETSI TIPHON Project. In progress documents at www.etsi.fr/tiphon. 1998.

[5] H. J. Fowler and J.W. Murphy. Network Management Considerations for Interworking ATM Networks with Non-ATM Services. IEEE Communication Magazine. Vol. 34, N.6, June 1996.

[6] C. Gbaguidi, J. P. Hubaux, G. Pacifici, and A. N. Tantawi. "An Architecture for the Integration of Internet and Telecommunication Services", Submitted to IEEE Journal. EPFL Technical Report SSC/1998/025. Also available at http://sscwww.epfl.ch

[7] M. Handley, H. Schulzrinne and E. Schooler. SIP: Session Initiation Protocol. Internet Draft. July 1997.

[8] IMTC- Voice over IP working group. www.imtc.org. 1998.

[9] ITU-T Recommendation E.164 (1997), The international public telecommunication numbering plan.

[10] ITU-T Recommendation G.723.1 (1996), Speech coders: Dual rate speech coder for multimedia communications transmitting at 5.3 and 6.3 kbit/s.

[11] ITU-T Recommendation H.323 (1998), Packet-based multimedia communications systems.

[12] ITU-T Recommendation H.225.0 (1998), Call signalling protocols and media stream packetization for packet based multimedia communication systems.

[13] ITU-T Recommendation H.245 (1998), Control protocol for multimedia communication.

[14] ITU-T Recommendation H.235 (1998), Security and encryption for H-Series (H.323 and other H.245 based) multimedia terminals.

[15] ITU-T Recommendation H.450.1 (1998), Generic functional protocol for the support of supplementary services in H.323.

[16] ITU-T Recommendation M.3010 (1996), Principles for a Telecommunications Management Network.

[17] ITU-T Recommendation Q.1211 (1993), Introduction to Intelligent Network Capability Set 1.

[18] ITU-T Recommendation Q.931 (1993), ISDN user-network interface layer 3 specification for basic call control.

[19] D. Klenke . Multimedia Management Information Base. Annex 2.4 - H.323 Gatekeeper. ITU-T Study Group 16. Q.12-14/16 Rapporteur Meeting Cannes, 8-11 June 1998.

[20] T. J. Kostas, M. S. Borella, I. Sidhu, G.M. Schuster, J. Grabiec and J. Mahler. Real-Time Voice over Packet Switched Networks. IEEE Network. Vol. 12, N. 1, Jan-Feb 1998.

[21] P. Mockapetris. Domain Names - Concepts and Facilities. IETF RFC 1034 . Nov. 1987.

[22] A.R. Modaressi and R.A. Skoog, Signaling System No 7: A tutorial. IEEE Communications Magazine. July 1990.

[23] B. Petri and D. Schwerje. Narrowband ISDN and Broadband ISDN Service and Network Interworking. IEEE Communication Magazine. Vol. 34, N.6, June 1996.

[24] J. Rosenberg and H. Schulzrinne. An RTP Payload Format for Generic Forward Error Correction. Internet Draft, IETF, Mar. 1998.

[25] J. Rosenberg and H. Schulzrinne. An RTP Payload Format for User Multiplexing. Internet Draft, IETF, Jan. 1998.

[26] H. Schulzrinne et al., RTP: A Transport Protocol for Real-Time Applications, IETF RFC 1889, Jan. 1996.

[27] L. Zhang, S. Deering, D. Estrin, S. Shenker and D. Zappala. RSVP: A New Resource ReSerVation Protocol. IEEE Network Magazine. 1993.