Computer Networks 53 (2009) 2050–2090

Contents lists available at ScienceDirect

Computer Networks

journal homepage: www.elsevier.com/locate/comnet

VoIP: A comprehensive survey on a promising technology

Stylianos Karapantazis *, Fotini-Niovi Pavlidou

Department of Electrical and Computer Engineering, Aristotle University of Thessaloniki, Panepistimioupoli, 54124 Thessaloniki, Greece

ARTICLE INFO

Article history:
Received 23 February 2008
Received in revised form 9 November 2008
Accepted 17 March 2009
Available online 28 March 2009

Responsible Editor: K. Kant

Keywords:
VoIP
IP Telephony
Voice quality
Voice codecs
Signaling protocols
Call admission control
Security

ABSTRACT

The Internet has burgeoned into a worldwide information superhighway during the past few years, giving rise to a host of new applications and services. Among them, Voice over IP (VoIP) is the most prominent one. Beginning more as a frolic among computer enthusiasts, VoIP has set off a feeding frenzy in both the industrial and scientific communities and has the potential to radically change telephone communications. In this article, we survey all these aspects that have the greatest impact on the quality of voice communications over IP networks. The survey begins with the merits and demerits of VoIP, followed by the Quality of Service (QoS) requirements that voice imposes and a description of test methods for the assessment of speech quality. We then proceed with a delineation of the issues related to the conversion of analog voice to packets, namely we spell out the details of the most well-known voice codecs, while light is also thrown on voice activity detection and voice packetization. Header compression schemes receive intense scrutiny as well. We also provide an overview of the signaling protocols that are tailored to the needs of VoIP, and we continue with the comparison of the call admission schemes that are geared towards the QoS constraints of VoIP. The pivotal issue of security is then discussed, pointing out potential threats as well as approaches for tackling them. Finally, the survey concludes with a discussion on the feasibility of providing VoIP over challenging satellite links.

1. Introduction

The Internet has burgeoned into a worldwide information superhighway during the past few years and wrought significant changes in the telecommunications arena. This unprecedented growth motivated the development of innovative applications with high bandwidth and low end-to-end delay requirements. One of the applications that thrived is Voice over IP (VoIP). VoIP, also known as IP or Internet telephony, is the technology that enables people to use the Internet as the transmission medium for voice communications. Beginning as a frolic among computer enthusiasts, VoIP has set off a feeding frenzy in both the industrial and scientific communities. Since its inception, huge strides have been made and now VoIP enjoys widespread popularity as an alternative to traditional telephony in homes and enterprises.

VoIP has the potential to revolutionize telephone communications. The trend toward voice communications over the Internet is mainly fuelled by the salient advantages Internet telephony offers. VoIP opens up exciting possibilities for users. In particular, it paved the way for monetary savings. It is cheaper for end-users to make an Internet telephone call than a circuit-switched call since most VoIP providers offer affordable long distance and international calling. Moreover, a significant amount of money can be saved on the monthly phone service considering that IP telephony service is less expensive than the traditional phone service. In addition, VoIP offers service flexibility since there are no dependencies between the application and the underlying network. VoIP users can already enjoy a variety of features, which they previously had to pay for, for free. Some of these features are voicemail, caller ID, call conferencing, call waiting and call forwarding. This

* Corresponding author. Tel.: +30 2310 996380.
E-mail addresses: skarap@auth.gr (S. Karapantazis), niovi@auth.gr (F.-N. Pavlidou).

1389-1286/$ - see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.comnet.2009.03.010

kind of users can also enjoy some new features like e-mailed voicemail and the easy management of contacts. Moreover, new services will commence to flourish as the market matures that will break the shackles of the Public Switched Telephone Network (PSTN) for voice services. One service that is expected to catch on is voice e-mail. Another key asset of IP telephony is flexibility. A user can move the IP Phone wherever he/she needs and still keep the same phone number. In addition, this technology offers a big step forward in available features and functions and allows end-users to control different media and different types of terminals from their Web browsers. Users will be able to set up conference calls from their homes, check the state of their accounts at any time and so on.

Several compelling reasons also impelled service providers to consider providing voice communications over the Internet. First and foremost, VoIP promises them new revenue sources. Further, new operators can capitalize upon this technology since it gives them an easy and cost-efficient way to compete with incumbent operators by undercutting their pricing regimes, while avoiding the regulatory hindrance to standard voice provision. From an engineering viewpoint, VoIP also holds considerable appeal because it gives carriers the ability to manage a single network that supports both voice and data.

Notwithstanding the aforementioned advantages, several issues must be accounted for while designing a VoIP system. VoIP is sometimes likened to mobile telephony because they both present some clear advantages over voice communications over PSTN. Mobile networks provide mobility and flexibility, whereas VoIP has an edge in cost savings. Moreover, VoIP performance is on a par with the performance of mobile cellular systems. On this account, there are some circumstances where the quality is considered acceptable. However, several strides remain to be made in order for VoIP to match the performance of the well-engineered PSTN. A laudable aim is to provide the same level of quality as PSTN does.

In this paper, rather than glossing over this promising technology, we aim to shed light to all these aspects that have the greatest bearing on voice quality (Fig. 1). We begin our survey with a reference to the Quality of Service (QoS) requirements that VoIP imposes, followed by a description of how certain performance metrics affect speech quality. Remedies are discussed as well. Before elaborating upon the factors that affect the performance of a VoIP system, we also spell out several test methods that have been devised to gage voice quality. Then, a plethora of voice codecs are delineated and their performance disparities are accentuated, along with all the aspects that should be contemplated before opting for a codec. Closely related to the selection of voice codecs are also the issues of silence suppression and voice packetization. Thus, we touch upon them as well. Another issue that receives intense scrutiny in our study is header compression techniques. Header compression techniques are the stepping stone to efficient bandwidth utilization. All the compression schemes that have been proposed thus far are compared against each other in order to come up with the most efficacious one.

Next we turn our attention to signaling protocols. Signaling protocols lay the foundations for high-quality voice communications since they are responsible for establishing and tearing down calls. We describe the salient features of all the well-known protocols with a view to listing their pros and cons. Moreover, we touch on the way Skype works, albeit information about Skype's signaling protocol is hard to procure. We also cast light to the way VoIP systems interface with PSTN. Next, we lay particular emphasis on an aspect that has often been considered as the downside of VoIP, that is, end-to-end QoS. Development of end-to-end QoS presents numerous challenges and is contingent upon the ability of the underlying network to provide differentiated QoS to various applications and users. Our attention is drawn towards call admission control (CAC) techniques. CAC and QoS are inextricably interwoven. A plurality of CAC techniques can be found in the literature. We describe all the schemes that are geared towards VoIP and some important conclusions are drawn. The next issue that we treat in this survey is security of VoIP networks. It is beyond the realms of possibility that a system will not experience malicious attacks that will aim to impair its performance. Contrary to the PSTN, VoIP systems are exposed to a number of threats. We first analyze the potential threats and then discuss several approaches on how to protect a VoIP network. We also dwell on the impact of security measures on QoS. Last but not least, we conclude this survey with a discussion on the feasibility of providing VoIP over satellite links. Next generation telecommunications systems can be viewed as an ultimate amalgamation of all existing and emerging fixed and mobile systems. In this context, satellite systems

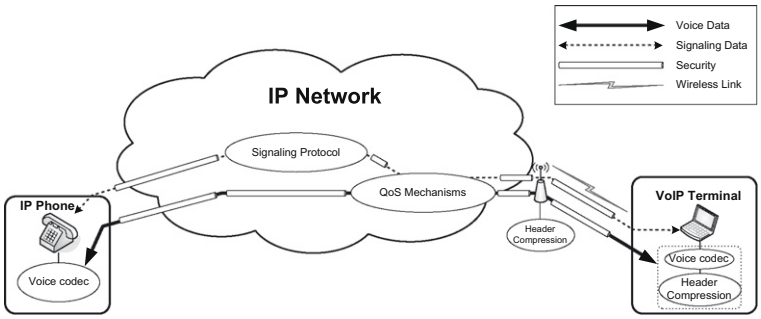

Fig. 1. General VoIP architecture.

will be called to support Internet telephony and we aim to cast light on the main problems that arise in this case.

2. QoS requirements of voice

Owing to advancement in technology over the years, traditional voice communication over the PSTN is characterized by high quality, often referred to as toll quality. When it comes to VoIP, stringent QoS constraints must be met in order to provide the same level of quality. Three are the main performance indicators that characterize the quality of voice communications over the Internet. The first one is average end-to-end delay, also referred to as one-way delay or mouth-to-ear delay. It is the time interval elapsed from the moment that the talker utters a word until the listener hears the word. The second metric is delay jitter, that is, the variation in the end-to-end delay between consecutive packets. Last but not least, the packet delivery ratio is of paramount importance as well, since voice is prone to packet loss.

2.1. Delay

As previously stated, VoIP is heavily impaired by long end-to-end delay. In general, there are five components that contribute to the mouth-to-ear delay:

(i) *Encoding delay:* the time interval needed to encode the voice signal, which depends on the voice codec employed.
(ii) *Packetization delay:* the interval that is required to packetize the encoded voice stream.
(iii) *Network delay:* the sum of transmission, propagation and queuing delays.
(iv) *Playback delay:* the delay induced by the playback buffer that resides at the receiver's side, which is needed to smooth delay jitter between consecutive packets.
(v) *Decoding delay:* the time interval needed to reconstruct the voice signal.

In some deployment scenarios, network delay may be increased by two factors. When one of the users that participates in a call resides in PSTN or in another IP network, then voice is subject to some delay at the interface between the VoIP system and PSTN or between the two IP networks. This delay is the result of the conversion of voice data from one format into another when a different voice codec is used in each network. This kind of delay, as well as remedies to it, are discussed in Section 4.4. Another component occurs when the user resides behind a residential Gateway.[1] In this case, packets accumulate at the gateway which then transmits them to the Internet. However, there exists some time intervals where the incoming traffic exceeds the capacity of the link that connects the gateway to the Internet. In that case, the packets are buffered before being transmitted and therefore, are subject to queuing delay.

Delay has two acute effects on voice performance. Firstly, it increases the subjective effect of any echo impairment. Secondly, as indicated in [1], even when echo is controlled, one-way delay affects the character of the conversation. Namely, for delays above 150 ms users will notice a slight hesitation in their partner's response, while for delays beyond 400 ms, the users should start to back off in order to preclude interruptions. ITU-T (International Telecommunication Union – Telecommunication Standardization Sector) Recommendation G.114 [1] provides some delay limits for one-way transmission on connections with adequate controlled echo. Table 1 tabulates these limits. In addition to the aforementioned categorization, the European Telecommunications Standard Institute's (ETSI) TIPHON project [2], which stands for Telecommunications and Internet protocol harmonisation over networks, defines five QoS classes and assigns a delay budget to each (Table 2). These two tables indicate that one-way delay should be less than 100 ms if toll quality is desired. However, there are also situations where longer delays and poorer quality are still acceptable, such as for access to remote or impervious regions.

A technique for dealing with the problem of delay is the prioritization of voice packets. Different scheduling techniques were compared in terms of queuing delay in [3]. Specifically, three schemes were examined, namely Priority Queuing, Weighted Round Robin and a scheme where voice is shielded from other traffic by having its own circuit. The analysis carried out in that study showed that Priority Queuing exhibits the best performance. More remedies for delay are discussed in Section 9.

2.2. Delay jitter

Delay jitter is the result of network congestion and improper queuing. At the sending side, voice packets are transmitted at a constant rate, while at the other end, packets may be received at an uneven rate. However, in order to reconstruct the initial analog voice stream the packets should be played out at a steady rate. If jitter is too large, that causes delayed packets to be cast aside, resulting in an audible gap. When a number of sequential packets are dropped voice becomes unintelligible. Human ear is highly intolerant of short-term audio gaps, therefore jitter should be kept to a minimum. Ideally, it should be less than 30 ms. However, depending on the delay budget, type of voice codec used and voice packet size, values between 30 and 75 ms can be acceptable as well [4].

The statistics of delay and delay jitter due to buffering of voice packets along with bursty background traffic were

Table 1
Delay limits for one-way transmission according to ITU-T Rec. G.114.

End-to-end delay (ms)	Quality
0–150	Acceptable for most users
150–400	Acceptable but has impact
400 and above	Unacceptable

[1] Residential gateway is called the device that connects a Local Area Network (LAN) to a Wide Area Network (WAN) or the Internet.

Table 2
ETSI TIPHON QoS classes.

QoS class	Wideband	Narrowband (High)	Narrowband (Medium)	Narrowband (Acceptable)	Best effort
End-to-end delay	<100 ms	<100 ms	<150 ms	<400 ms	<400 ms

evaluated in [5,6]. Using a theoretical analysis, the authors of those studies assessed the impact of specific characteristics of background traffic, such as traffic load, burstiness and burst length on delay jitter. An increase in any of these three parameters exacerbates voice quality.

In VoIP implementations, the mechanism that is used to compensate for delay jitter and smooth out the distribution of packet delay is the play-out buffer, also referred to as jitter buffer. Its function is to buffer packets and then play them out at a steady rate. Jitter buffers are categorized into static buffers and dynamic buffers [7]. The former type of buffers is the easiest to implement, whereas their size is fixed. At the other extreme, dynamic buffers can adjust the play-out point in the buffer based on the history of delay jitter. Nonetheless, it is hard to determine the right time to adjust the play-out point. One important consideration that should not be overshadowed is that of the relation between the delay added by a jitter buffer and the number of packets that it drops. On the one hand, longer delay translates into a decrease in the number of dropped packets, however, it reduces speech quality. On the other hand, if the delay that the buffer adds is decreased, then inevitably the number of the packets that are cast aside increases. Therefore, there is a trade-off that must be sought between these two metrics.

2.3. Packet loss

Packet loss represents another meaningful metric. It has been reckoned that delivery ratio of more than 99% is required for VoIP [8–10]. Packet losses can occur due to many reasons. An IP packet may be lost due to congestion in the IP network, i.e., a router is deluged with a raft of packets it cannot handle. An IP packet may also be discarded at the destination, for example, when a packet arrives too late to be played out. In addition to these reasons, in wireless systems intense variations in the received signal power, brought about by intense fading conditions, result in the loss of a significant number of packets too. The latter is perceived by end-users as voice clipping. Wireless links present a challenge, even without VoIP. VoIP compounds the situation even further. VoIP is quite robust in that it can tolerate bit error rates (BERs) as high as 10^{-5}, but for higher rates voice becomes unintelligible. Bit errors may corrupt signaling packets, causing failure of call establishment, or media packets, resulting in degradation of voice quality. Fading may corrupt IP headers or the packet's payload. In the former case, packets are dropped by routers, whereas in the latter case, the payload cannot be successfully decoded, hence resulting in voice frame loss. Typical guidelines for packet loss rates are 1% or less for toll quality and 3% or less for business quality [8]. A figure that relates speech quality to packet loss can be found in [7]. The impact of packet loss on voice quality as perceived by VoIP users was investigated in [11] as well. It was concluded that voice quality is acceptable only if packet loss is less than 2%.

To alleviate the detrimental effects of packet loss due to channel impairments several forward error correction (FEC) techniques have been proposed with the aim of diminishing the number of errors in data transmission [12,13]. FEC is implemented by adding redundancy to the transmitted information using a predefined algorithm, thereby allowing the receiver to detect and correct errors within a bound. However, the benefits of FEC are partly negated by the increase in delay and bitrate. The audio quality under a simple FEC scheme was evaluated in [14]. Specifically, the FEC scheme under consideration adds some redundant information of packet n to the next packet n + 1, so that if packet n is dropped, it can be recovered if packet $n + 1$ is correctly received. The redundant information about packet n is obtained using a low bitrate encoder. However, ample simulation results led to the conclusion that this FEC scheme is not apt for improving voice quality when a large number of flows employ similar FEC techniques. Pursuing an optimal FEC scheme for VoIP, the authors of [15] developed a scheme that adapts itself to varying loss conditions in order to provide acceptable voice quality with the minimum redundancy possible. In that FEC scheme, each packet carries redundant information about the previous K packets, where K is a parameter of the algorithm and can be dynamically changed. The encoding rate that is employed for each copy of a packet should, nevertheless, be carefully opted. Preliminary simulation results confirmed the positive characteristics of that scheme. A technique akin to this FEC scheme was also described in [16]. The difference between these two schemes lies in the procedure that is used to determine how much of redundant information a packet should carry. In addition to FEC techniques, some voice codecs employ their own error concealment methods, such as packet replication or silence substitution. However, these techniques are most effective for low packet loss rates around 2%. In [17] some error concealment strategies are delineated, along with two new techniques that are based on linear prediction to replace or correct damaged packets. For details on packet loss recovery techniques in general, the reader is referred to [18].

3. Quality assessment

Since the current Internet was not designed to transport voice traffic, an important aspect in VoIP communications is the assessment of voice quality. It is imperative that new voice services undergo a significant amount of testing to evaluate their performance. In this section we outline the test methods that are geared towards the evaluation of VoIP systems.

Testing methods can be classified into subjective and objective tests [7,11]. Subjective methods rely on the opinion of a panel of listeners, who are usually asked to rate the quality of the test sentences read aloud by both male and female speakers[2] over the communications medium being tested. Then a Mean Opinion Score (MOS) is computed by averaging all votes out. MOS is expressed as a single number on a scale from 1 to 5, where 1 represents the lowest perceived quality, while 5 is the highest perceived quality. MOS tests for voice are specified in ITU-T Recommendation P.800 [19]. The performance of the system under test can be rated either directly (i.e., absolute category rating, ACR) or relative to the subjective quality of a reference system (i.e., degradation category rating, DCR). An overview of subjective quality assessment methodologies can be found in [20]. However, the validity of the results obtained with subjective testing can be questioned. Subjective tests suffer from a number of problems and limitations. First of all, it takes a long time to perform these tests, while the lack of a "standard" environment for listening to take place may also affect reproducibility or the ability of the test to predict user satisfaction at home. Moreover, subjective testing is ill-suited for examining the impact of a large number of parameters on voice quality. On the other hand, this type of evaluation is easy and inexpensive to perform.

Objective testing methods are based on measurements of physical quantities of the system such as delay, delay jitter and packet loss. Typically, this can be achieved either by injecting a test signal into the system or by monitoring live traffic. Contrary to subjective tests, objectives tests can be repeatedly carried out to evaluate the performance of a system under different set of parameters. This type of tests can be further divided into intrusive and non-intrusive [7]. In the former kind of tests, the system is taken out of service and tested by the insertion of a test signal. In non-intrusive tests, measurements are performed on live user traffic. The latter method of testing allows for a larger number of tests without any loss of revenue. The pitfall, however, is that these measurements are hard to develop and are, in general, less accurate than intrusive methods. A description of the most well-known objective testing methods follows.

3.1. PSQM

The Perceptual Speech Quality Measure (PSQM) represents the first objective testing method developed by KPN in 1996. It is defined in ITU-T Recommendation P.861 [21]. At the time that this method was standardized, the scope was to devise a method for evaluating voice codecs used primarily in cellular systems. The measure of quality predicted by PSQM is given on a scale from 0 to 6.5. PSQM is appropriate for evaluating speech quality in environments that are not subject to bit or frame errors. Thus, this standard is not suited for the assessment of networks, but rather for codecs, and it was withdrawn when PESQ was adopted in February 2001.

[2] Telephony systems are reckoned to work better in passing some voices than others [7].

3.2. PAMS

The Perceptual Analysis Measurement System (PAMS) was developed by British Telecom in order to evaluate the perceived voice quality [22]. PAMS was the first method to provide robust assessment for VoIP. This repeatable testing method derives a set of scores by comparing one or more high-quality reference speech samples to the processed audio signal. The resulting score is on a MOS-like scale from 1 to 5. Nevertheless, PAMS does not always show 100% correlation with live-listener MOS test scores conducted on the same speech samples.

3.3. PESQ

The Perceptual Evaluation of Speech Quality (PESQ) method is an intrusive method developed jointly by British Telecom and KPN. The details of this method are spelt out in ITU-T Recommendation P.862 [23]. PESQ combines the excellent psycho-acoustic and cognitive model of PSQM with a time alignment algorithm adopted from PAMS that is able to handle varying delays perfectly. In this method, the received signal is compared to the one initially transmitted using a perceptive hearing model that is a replica of the human hearing apparatus. The result of this comparison is given on a scale from 1 to 4.5.

3.4. E-Model

A useful tool for assessing the relative impact of transmission planning decisions on speech performance is the E-Model. The E-Model has been included in various ITU-T Recommendations on transmission planning. In particular, the E-Model is the subject of Recommendation G.107 [24]. Notwithstanding, it has also been adopted by ETSI and TIA (Telecommunications and Industry Association) and has become the most widely used tool for objective assessment of speech quality. This model is predicated upon the assumption that the impairments caused by transmission parameters have an additive effect on speech quality. According to this model, speech quality is determined by the following equation:

$$R = R_0 - I_s - I_d - I_e + A, \tag{1}$$

where R_0 accounts for noise effects, I_s represents impairments such as too loud speech level, non-optimum sidetone and quantization noise, I_d is the sum of impairments due to delay and echo effects, I_e represents impairments due to low bitrate voice codecs, whereas A represents an "advantage of access" that some systems have in comparison with PSTN (for instance, A for mobile systems is 10). It should be noted that the first two terms in Eq. (1) are intrinsic to voice signal itself and do not depend on the transmission of voice over the Internet. The value of this function has a nominal range from 0 for terrible up to 100 for perfect voice. However, there is a direct relation between R and the MOS score. Fig. 2 depicts the relation among PSQM, PAMS, PESQ, MOS and E-Model scales. Even though the E-Model is widely used for the assessment of voice quality, it can also be used for opting some network

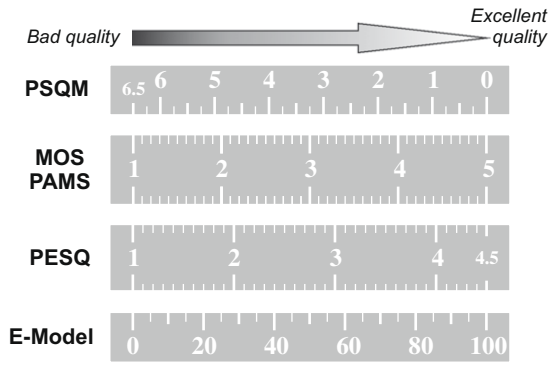

Fig. 2. PSQM, PAMS, PESQ and MOS scales in comparison.

parameters such as the voice codec and the maximum allowable link utilization [25].

3.5. P.563

Recently, ITU ratified a new non-intrusive voice quality test method. This method, which is specified in [26] represents the fruits of the collaborative efforts of three companies, namely Psytechnics, OPTICOM and SwissQual. Contrary to PSQM and PESQ, this method is non-intrusive, therefore a test signal is not required. This type of testing can be applied all the time for monitoring network quality. However, under critical flaws, an intrusive method ought to be used since it provides more accurate results.

4. Voice codecs

Since the early days of networking bandwidth has been considered a resource at a premium. Therefore, significant efforts have been drawn towards the minimization of the amount of bandwidth required by specific services in order for the network to be able to serve a greater number of users. In this context, compressing voice signals while keeping the quality perceived by users at acceptable levels represents a daunting challenge. This section is devoted to the methods that either are currently in use or have been proposed for the compression of audio signals, which are referred to as voice codecs.

Voice codecs are the algorithms that enable the system to carry analog voice over digital lines. There are several codecs, varying in complexity, bandwidth needed and voice quality. The more bandwidth a codec requires, normally the better voice quality is. One problem that arises in the delivery of high-quality speech is network efficiency. Albeit it is feasible to provide high-quality speech, this comes at the expense of low network efficiency. Most domestic PSTN operate with voice sampled at 8 kHz and an 8-bit non-linear quantization scheme according to ITU-T G.711 [27], which encodes at 64 kb/s. Nonetheless, a much lower bitrate is desirable for a number of applications on account of the limited capacity or in order to maximize the amount of traffic that can be carried over the network. There exist several codecs that compress voice to as low as 2.15 kb/s, quality is well below that of G.711 though. Next, we summarize the most well-known voice codecs, which can also be found in the bulk of VoIP solutions. For the sake of presentation, the codecs are classified into three categories. Namely, narrowband codecs that operate on audio signals filtered to a frequency range from 300 to 3400 Hz and sampled at 8 kHz, broadband codecs that operate on audio signals filtered to a frequency range from 50 to 7000 Hz, and sampled at 16 kHz and multimode codecs that can operate on either narrowband or broadband signals.

4.1. Narrowband codecs

G.711

G.711 is a Pulse Code Modulation (PCM) scheme that produces one 8-bit value every 125 μs, resulting in a 64 kb/s bitstream [27]. Each audio data is encoded as eight bits after logarithmic scaling. This standard has two forms, μ-Law (used in North America and Japan) and A-Law (used in Europe and the rest of the world). An A-Law G.711 PCM encoder converts 13 bit long linear PCM samples into 8 bit compressed PCM (logarithmic form) samples, and the decoder does the conversion vice versa, whilst an μ-Law G.711 PCM encoder converts 14 bit linear PCM samples into 8 bit compressed PCM samples. The G.711 is the standard codec used in H.323 and the Integrated Services Digital Network (ISDN) network.

G.723.1

G.723.1 is a dual rate speech codec standard from ITU-T, originally developed for videophones that deliver video and speech over regular phone lines (PSTN) [28]. It was designed for the ITU-T H.323 and H.324 audio and videoconferencing/telephony standards for compressing the toll quality speech. G.723.1 was standardized in 1996 as a part of the overall H.324 family of standards and can operate at two bit rates:

- 6.3 kb/s (using 24 byte chunks) using a Multi Pulse-Maximum Likelihood Quantization (MPC-MLQ) algorithm.
- 5.3 kb/s (using 20 byte chunks) using an Algebraic Code Excited Linear Prediction (ACELP) algorithm.

The implementation of G.723.1 Annex A also includes silence compression techniques to reduce the transmitted bitrate during the silent intervals of speech. The additional advantages that accrue from the use of Voice Activity Detection (VAD) consist in using lower processing loads and bandwidth during silence intervals.

G.726

ITU-T G.726 superseded ITU-T G.723 [29]. It works at four bitrates, i.e., 16, 24, 32 and 40 kb/s [30]. Specifically, this codec is recommended for the conversion of a single 64 kb/s A-law or μ-law PCM channel encoded at 8 kHz to a 16, 24, 32 or 40 kb/s channel. It works based on the principle of ADPCM. Nonetheless, the 16 and 24 kb/s encoding rates do not provide toll quality speech. Therefore, ITU-T G.726 recommends that the 16 and 24 kb/s rates should be alternated with higher data rate encoding to provide an average sample size of between 3.5 and 3.7 bits per sample.

G.728

ITU-T G.728 describes a low delay speech codec for the coding of speech signals at 16 kb/s using low delay code excited linear prediction (LD-CELP) [31]. G.728 Annex G (G.728 G) is a fixed point specification of the codec working at a bitrate of 16 kb/s. G.728 Annex I (G.728 I) is the packet loss concealment (PLC) technique used along with G.728 codec. This is a very robust speech codec, with very good speech quality, comparable to 32 kb/s ADPCM.

G.729/G.729 Annex A

The G.729 codec allows for stuffing more calls in limited bandwidth in order to utilize IP voice in more cost-effective ways [32]. The basic algorithm of G.729 runs at 8 kb/s and is optimized to represent speech with high quality. It uses the conjugate structure algebraic code excited linear prediction (CS-ACELP) algorithm with 10 ms frames. However, the complexity of this algorithm is rather high. To this end, ITU-T came up with the G.729 Annex A (G.729A) codec [33], which is a medium complexity variant of G.729 with slightly lower voice quality. The G.729 and G.729A codecs are inter-operable, i.e., speech coded with G.729 can be decoded by G.729A decoder and vice versa. G.729A is a very robust speech codec that works at bitrate of 8 kb/s with very good speech quality comparable to 32 kb/s ADPCM. Like G.729, G.729A is also based on the principle of CS-ACELP. The Annex B of G.729 adds functionality to the G.729 family of codecs [34]. In essence, it comprises a VAD module, a DTX (Discontinuous Transmission) module which decides on updating the background noise parameters during silence periods, as well as a CNG (Comfort Noise Generation) module.

G.729 Annex D

Annex D of the G.729 recommendation was approved in September 1998 [35]. This annex describes a lower bitrate codec that operates at 6.4 kb/s. This codec can be employed during periods of congestion, so that operation can continue at 6.4 kb/s with minimal degradation of speech quality, or when more bits are needed by the forward error correction scheme to compensate for channel impairments.

G.729 Annex E

Annex E of the G.729 recommendation was also approved in September 1998 [36]. This annex delineates a higher bitrate codec that can be used when bandwidth is available in order to improve performance in the presence of background noise and music. G.729E describes a codec operating at 11.8 kb/s that encodes each frame in two different ways and then selects the method that appears to provide the greatest fidelity. The difference between the two methods lies in the algorithm used for the compression. One is based on CS-ACELP, whereas the other features a backward-adaptive Linear Predictive Coding (LPC) synthesis filter.

GSM-FR

The ETSI GSM 06.10 Full Rate (FR) codec was the first digital speech coding standard used in GSM (Global System for Mobile Communications) digital mobile phone systems, working on an average bitrate of 13 kb/s [37]. Introduced in 1987, this codec uses the principle of Regular Pulse Excitation-Long Term Prediction-Linear Predictive (RPE-LTP) coding. The speech encoder takes its input as a 13 bit uniform PCM signal either from the audio part of the mobile station or, on the network side, from the PSTN via an 8 bit/A-law to 13 bit uniform PCM conversion. The quality of the encoded speech is quite poor in modern standards, but at the time of its development it was a good compromise between computational complexity and quality. The codec is still widely used in networks around the world.

GSM-HR

The GSM 06.20 GSM half rate (HR) codec was introduced in 1994 for use in GSM. It uses the VSELP (Vector-Sum Excited Linear Prediction) algorithm, which translates into a greater need for processing power [38]. GSM-HR's average bitrate is 5.6 kb/s. Since the codec, operating at 5.6 kb/s, requires half the bandwidth of the full rate codec, network capacity for voice traffic is doubled, at the expense of lower audio quality.

GSM-EFR

The GSM 06.60 Enhanced Full Rate (EFR) codec was introduced in 1997 and constitutes an improvement on GSM-FR [39]. It is based on ACELP and its bitrate is 12.2 kb/s. Albeit it consumes less bandwidth, it provides better speech quality and is generally more robust in regard to network impairments.

MELPe

The MELPe (Mixed-Excitation Linear Predictive) Vocoder Algorithm is the new 1.2 and 2.4 kb/s Federal Standard speech codec selected by the United States Department of Defence (DoD) Digital Voice Processing Consortium (DDVPC) after a multi-year extensive testing program. Its frame interval is 22.5 ms ± 0.01 percent in duration and contains 180 voice samples (8 kHz). MELPe is robust in difficult background noise environments such as those frequently encountered in commercial and military communication systems. It is very efficient in its computational requirements. This translates into relatively low power consumption, an important consideration for portable systems.

AMR

The Adaptive Multi-Rate (AMR) speech codec standard was introduced by the 3rd Generation Partnership Project (3GPP) for compressing toll quality speech (8 kHz) [40]. This speech codec was designed with the volatility of the wireless medium in mind for speech compression in the 3rd generation (3G) mobile telephony. This codec operates at eight basic bitrates, 12.2, 10.2, 7.95, 7.40, 6.70, 5.90, 5.15 and 4.75 kb/s, allowing the on-the-fly switch between different rates. It uses the principle of ACELP for all bitrates. Besides, there are two types of VAD and CNG algorithms. Moreover, it was specifically designed to improve link robustness. AMR supports dynamic adaptation to network conditions, using lower bitrates during network congestion, while preserving audio quality at an acceptable level. By trading off the speech bitrate to channel coding, AMR maximizes the likelihood of receiving the signal at the far end. AMR can be considered to be the most widely deployed codec in the world today.

iLBC

iLBC stands for Internet Low Bitrate Codec and is a royalty-free narrowband speech codec, developed by Global IP Sound (GIPS) [41,42]. The fact of being freeware led to the adoption of iLBC in many commercial and free applications

such as Skype, the Gizmo Project, OpenWengo and Google Talk. It has support for two basic frame lengths: 20 ms at 15.2 kb/s and 30 ms at 13.33 kb/s. When the codec operates at block lengths of 20 ms, it produces 304 bits per block. Similarly, for block lengths of 30 ms it produces 400 bits per block. Further, this coded uses a block-independent LPC algorithm. The fact of encoding each block of samples independently of the previous ones makes this codec able to withstand a certain degree of frame losses. Notwithstanding, while this provides better quality when 10% (or more) of the packets are being dropped, this makes the codec suboptimal for clean line conditions.

4.2. Broadband codecs

G.722

G.722 was the first wideband voice codec standardized by ITU-T, mainly for use in ISDN videoconferencing [43]. G.722 can handle speech and audio signals of bandwidth up to 7 kHz, compared to 3.6 kHz in the case of narrowband speech codecs. G.722 relies on the principle of Sub-Band – Adaptive Differential Pulse Code Modulation (SB-ADPCM). The signal is split into two sub-bands and samples from both bands are coded using ADPCM techniques. The system has three basic modes of operation corresponding to the bitrates used for 7 kHz audio coding: 64, 56 and 48 kb/s. The latter two modes allow an auxiliary data channel of 8 and 16 kb/s, respectively, to be provided within the 64 kb/s by making use of bits from the lower sub-band.

G.722.1

G.722.1 is an ITU-T standard audio codec meant for low bitrate audio coding below 64 kb/s, namely 24 kb/s or 32 kb/s [44]. The algorithm is based on transform technology, using a Modulated Lapped Transform (MLT), and operates on 20 ms frames of audio. Because the transform window is 640 samples and a 50% overlap is used between frames, the effective look-ahead[3] buffer size is 20 ms.

G.722.2

Using technology developed jointly by VoiceAge and Nokia, G.722.2 is the first codec to be adopted as a standard for both wireless and wireline services. It was standardized first by ETSI/3GPP in December 2001 as AMR-WB (Adaptive Multi-Rate Wideband) and then approved by ITU-T in January 2002 as the G.722.2 recommendation [45]. The codec supports nine different bitrates for the encoded stream: 6.60, 8.85, 12.65, 14.25, 15.85, 18.25, 19.85, 23.05 or 23.85 kb/s. The coding scheme for the multi-rate coding modes is ACELP.

AMR-WB+

AMR-WB+ is an audio codec that extends AMR-WB (that is, ITU-T G.722.2 codec) [46]. It adds support for stereo signals and higher sampling rates. Another main improvement is the use of transform coding additionally to ACELP. This greatly improves the generic audio coding.

G.729.1

G.729 has been extended to provide support for wideband speech and audio coding, i.e., the transmitted acoustic frequency range is extended to 50 Hz–7 kHz [47]. The respective extension to G.729 is referred to as G.729.1. Nonetheless, the ITU-T G.729.1 codec can also be used as a narrowband codec to encode 50–4000 Hz signals at bitrates of 8 and 12 kb/s (its frame size is 20 ms). Thus, it can be used in existing G.729 based VoIP systems.

iSAC

Except for iLBC, Global IP Sound is also the developer of iSAC (internet Speech Audio Codec), which is used by Skype, the Gizmo Project and Google Talk. However, unlike iLBC, this codec is proprietary. iSAC is a wideband speech codec that delivers better than PSTN sound quality. To attain the best quality for the existing connection speed, it adjusts the bitrate from 10 kb/s up to 32 kb/s, thereby achieving a good trade-off between bandwidth and speech quality. This codec is prone to packet loss and is also available in a low-complexity mode for usage in PDAs and mobile phones.

4.3. Multimode codecs

Speex

Speex is an Open Source/Free Software patent-free audio compression format designed for speech [48,49], commenced as a one-project in February 2002. Speex is well-suited to handle VoIP, internet audio streaming, data archival (like voice mail) and audio books. Currently, LinPhone, OpenH323 and GnomeMeeting are some of the projects using Speex. Speex is mainly designed for three different sampling rates: 8, 16 and 32 kHz. These are respectively referred to as narrowband, wideband and ultra-wideband. It is based on CELP and is designed to compress voice at bitrates ranging from 2.2 to 44 kb/s. It is also one of the few codecs that can change bitrate dynamically, at any moment. In Speex, sounds like vowels and high-energy transients require higher bitrate to achieve good quality, while fricatives (e.g., s and f sounds) can be coded adequately with fewer bits. For this reason, variable bitrate (VBR) can achieve lower bitrate for the same quality, or a better quality for a certain bitrate. Speex also employs packet loss concealment, which makes it an interesting candidate for the wireless environment. Despite its advantages, VBR has two main drawbacks: first, by only specifying quality, there is no guarantee about the final average bitrate. Second, for some real-time applications like VoIP, what matters is the maximum bitrate, which must be low enough for the communication channel.

BroadVoice

BroadVoice is a family of voice codecs developed by Broadcom for VoIP applications such as Voice over Cable, Voice over DSL (half the bandwidth) and IP phone applications [50,51]. The BroadVoice codec family consists of two codec versions. The narrowband version of BroadVoice, called BroadVoice16 (BV16) and the wideband version termed BroadVoice32. As it becomes evident from the names given to these two versions, the first one operates

[3] Look-ahead is the period of time when the codec waits for the first part of the following frame for patterns that it can compress while coding the current packet. In particular, the algorithm must have some knowledge of what is in block $N+1$ in order to accurately reproduce sample block N at the decoder side. This look-ahead, which is in fact an additional delay, is also called algorithmic delay.

at 16 kb/s (sampling rate is 8 kHz), while the second at 32 kb/s (sampling rate is 16 kHz).

4.4. Codec selection

The selection of a codec is a matter of paramount importance and profoundly affects the system performance. As can be seen in Table 3, which tabulates the most important characteristics of the most well-known voice codecs, there exist significant differences among the codecs, and what is more, several aspects should be taken into consideration before opting for the most appropriate codec. As far as codec delay is concerned, it must be pointed out that according to ITU, it is twice the frame size plus any potential look-ahead time. In other words, the processing time is equal to the frame size. The rationale behind the previous statement is that the codec should complete the processing of the samples and produce a stream of bits before the next group of samples becomes available. The ITU definition of codec delay gives, however, the worst-case results. Given the current state-of-the-art in Digital Signal Processors (DSPs), more processing power can become available for voice compression, which in turn will result in a reduction in the codec delay.

It has been reckoned thus far that the choice of codec is to some extent dictated by the bandwidth on offer which determines the maximum bitrate for the codec, and in turn the maximum voice quality that the system can attain under ideal conditions. Given the demand for high quality, it becomes apparent that choosing the lowest bitrate codec will not suit a large proportion of today's VoIP market. To this end, most VoIP systems tend to support G.711 and at least one low bitrate codec to allow the operator some flexibility to trade between quality and bandwidth, potentially on a per call basis [7]. However, when designing a VoIP system it is important to consider how well it will work with existing networks. Since there exists a plurality of voice codecs, it is most likely that different systems will make use of different codecs. In that case, at the point where the networks join, voice should be decoded and re-encoded. This results in a transcoding or tandeming of speech codecs. Different codecs react differently to tandem encoding and generally the decompression/compression cycle exacerbates voice quality [52,53]. The order of codecs along the end-to-end path impacts on the quality since speech is distorted in a non-linear way. For example, G.729 followed by GSM-EFR will not produce the same quality as GSM-EFR followed by G.729 [7]. Further, the concatenation of voice codecs significantly increases delay due to the time interval needed for decoding and encoding. A solution to this problem is the use of a common codec end-to-end. The choice of the codec can be negotiated at call setup time

Table 3
Characteristics of the most well-known voice codecs.

Codec	Bitrate (kb/s)	Frame (ms)	Bits per frame	Algorithmic delay[a] (ms)	Codec delay[b] (ms)	Compression type	Complexity (MIPS)[c]	MOS
Narrowband codecs								
G.711	64	0.125	8	0.125	0.25	PCM	$\ll 1$	4.1[d]
G.723.1	6.3	30	189	37.5	67.5	MP-MLQ	$\leqslant 18$	3.8
G.723.1	5.3	30	159	37.5	67.5	ACELP	$\leqslant 18$	3.6
G.726	16	0.125	2	0.125	0.25	ADPCM	≈ 1	–
G.726	24	0.125	3	0.125	0.25	ADPCM	≈ 1	3.5
G.726	32	0.125	4	0.125	0.25	ADPCM	≈ 1	4.1
G.728	16	0.625	10	0.625	1.25	LD-CELP	≈ 30	3.61
G.729	8	10	80	15	25	CS-ACELP	$\leqslant 20$	3.92
G.729A	8	10	80	15	25	CS-ACELP	$\leqslant 11$	3.7
G.729D	6.4	10	64	15	25	CS-ACELP	<20	3.8
G.729E	11.8	10	118	15	25	CS-ACELP LPC	<30	4
GSM-FR	13	20	260	20	40	RPE-LTP	≈ 4.5	3.6
GSM-HR	5.6	20	112	24.4	44.4	VSELP	≈ 30	3.5
GSM-EFR	12.2	20	244	20	40	ACELP	≈ 20	4.1
AMR-NB	4.75–12.2	20	95–244	25	45	ACELP	15–20	3.5–4.1
iLBC	13.33	30	400	40	60	LPC	18	3.8
iLBC	15.2	20	304	25	40	LPC	15	3.9
Speex (NB)	2.15–24.6	20	43–492	30	50	CELP	8–25	2.8–4.2
BV16	16	5	80	5	10	TSNFC	12	4
Broadband codecs								
G.722	48, 56, 64	0.0625	3–4	1.5	1.5625	SB-ADPCM	5	~ 4.1
G.722.1	24,32	20	480, 640	40	60	MLT	<15	~ 4
AMR-WB (G.722.2)	6.6-23.85	20	132-477	25	45	ACELP	≈ 38	Various
Speex (WB)	4-44.2	20	80-884	34	50	CELP	8-25	Various
iSAC	Variable 10–32	Adaptive 30–60 ms	Adaptive-variable	Frame + 3 ms	Adaptive 63–123	Transform coding	6–10	Various[e]
BV32	32	5	160	5	10	TSNFC	17.5	~ 4.1

[a] Every speech codec introduces a delay in the transmission. This delay amounts to the frame size, plus some amount of "look-ahead" required for examining future samples.
[b] The codec delay is the sum of the algorithmic delay and the time interval needed for processing purposes.
[c] MIPS stands for million instructions per second and represents a measure of a computer's processor speed.
[d] Theoretical maximum: 4.4.
[e] Better than G.722.2 at comparable bitrates.

between the involved networks. Notwithstanding the benefits stemming from this approach, this solution poses several new engineering challenges from the networking viewpoint, the most important one being the efficient bandwidth utilization. While the bitrate of some codecs may be acceptable in some networks, there may exist several scenarios where their use may be prohibited. For instance, the use of G.711, which is used in PSTN, in wireless networks would result in a significant decrease in the number of simultaneous calls that the system can support. Therefore, the codec negotiation procedure should take account of the kind of network where each terminal resides as well as its level of congestion. However, the whole procedure should not increase call setup delay unacceptably, while its complexity should not be cumbersome.

Another important characteristic that should be taken into account is the constraints imposed by the network itself. For instance, satellite networks suffer from their intrinsic high propagation delay, especially in the case of Geostationary (GEO) systems where propagation delay is around 250 ms. That means that if one-way delay must be between 300 and 400 ms, then all the other tasks related to the voice compression and digitization of analog voice should be completed within a time window equal to 50–150 ms. Hence, in such a case the codec delay becomes an important criterion for the decision on the most appropriate codec. In addition, the choice of the codec may also be dictated by the number of bits per frame. In many systems, access to the network is divided into fixed-length slots. Thereby, the codec to be selected must be suitable for transmission onto these slots.

Last but not least, another important issue that arises is the cost associated with each codec. Some codecs such as the majority of the G.7xx codecs family are not free to use, thus a fee must be paid to use them. Nonetheless, the last few years have witnessed the proliferation of open-source solutions such as Speex and BroadVoice that are royalty and license free.[4] Aside from reducing the overall cost of the network (for example, the licensing fee for the G.729A codec are $15,000 initial fee plus $1.45–$0.30 per channel depending on total volume [54]), the effort of coping with the complicated, multi-party intellectual property rights is also avoided.

All the aforementioned statements lead to the conclusion that the selection of a codec is not a trivial matter. There does not exist one codec that outperforms the rest of voice codecs in any scenario. While some codecs seem suitable for some cases, they may be ill-suited to other scenarios. Hence several factors should be taken into account before opting for a voice codec.

5. Voice Activity Detection

The Voice Activity Detection (VAD) technique stops sending data when there is no voice signal. This can normally achieve a 30–35% saving of the bandwidth, and can be implemented either directly into the codec or at the voice concentrator. However, the variety and the varying nature of speech and background noise makes VAD a daunting challenge. Real speech is distorted by background noise, especially in distant-talking situations. Inaccurate detection of the speech periods causes deterioration of speech quality. One of the problems that arise is the susceptibility of silence suppression to front and back-end clipping [7]. In other words, there are times when VAD is triggered either too early or too late, thereby excising speech from the beginning or the end of a sentence. A comfort noise generator that allows the insertion of an artificial noise during silent intervals of speech is also used in conjunction with VAD, because in a communication channel, if transmission is stopped because there is no speech, then the user may assume that link has been cut. Nevertheless, if a mismatch in noise between active and silent periods occurs, then the result is an impression of poor quality.

6. Packetization efficiency

Related to the problem of choosing the most appropriate codec is also the question of how much speech can be placed in an IP packet. Voice is loss-tolerant and delay sensitive, therefore voice is carried over the User Datagram Protocol (UDP). In addition, since VoIP applications are real-time, the Real-Time Transport Protocol (RTP) runs on top of UDP [55]. On the one hand, sending small packets is inefficient due to the increased header overhead. On the other hand, large packets lead to long end-to-end delay, which is inapt for real-time communications. The impact of packet size on speech quality was evaluated for G.711 and G.729 under various network conditions in [56]. The experiments conducted in that study shown that for very low packet loss rates, speech quality is not affected by the size of the packet. Nonetheless, for moderate and high packet loss rates, long packets result in higher deterioration in voice quality in comparison with short packets. This trend can easily be explained bearing in mind that a long packet carries a larger number of voice frames. Moreover, each codec exhibits different levels of tolerance to packet loss. In addition, as mentioned earlier, some systems operate with packets of fixed length. Therefore, the codec's frame size plays an important role on this decision.

However, it should be noted that due to the IP, UDP and RTP headers which contribute towards a 40-byte overhead to each packet, it makes limited sense to select a codec whose rate is below a certain level. This becomes more obvious from Fig. 3 which depicts the efficiency gain from using lower bitrate codecs (10 and 20 ms packet spacing is assumed). Codec efficiency is defined as the reduction in bitrate that a codec achieves compared to the 64 kb/s codec (G.711). This figure does not aim to compare codec schemes against each other. Rather, it intends to accentuate the fact that there is a certain level of compression under which there does not exist significant gain in bandwidth savings when the IP/UDP/RTP headers are attached to each packet. The benefit stemming from the employment of a low bitrate codec lies in bandwidth saving. The price to pay is, however, a lower perceived quality of voice. Fig. 2 illustrates that the bandwidth saving compared to the 64 kb/s codec used in PSTN networks is re-

[4] iLBC is also royalty free, though not open-source.

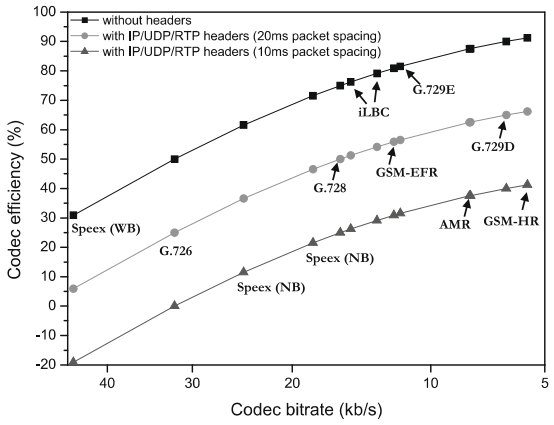

Fig. 3. Codec efficiency with and without IP/UDP/RTP headers.

duced when the IP/UDP/RTP headers are taken into account. Therefore, it makes limited sense to compress voice beyond a certain level since when accounting for the IP headers the bandwidth saving is reduced. Moreover, as packet spacing decreases, so does delay, but overhead increases. The authors of [3] studied the effect of packet size on voice delay and bandwidth utilization for three voice codecs: G.711, G.723.1 and G.729A. It was shown that a packet spacing of 30 ms provides a reasonable trade-off between low delay and efficient network utilization with regard to G.729.A and G.723.1, while for G.711 the optimal packet formation time is 10 ms.

7. Header compression

As mentioned earlier, in applications such as VoIP the payload of the IP packet is roughly of the same size or even smaller than the headers. Even though these headers are important over the end-to-end path, they serve no significant purpose over just one link. Thus, it is possible to compress them in order to utilize bandwidth more efficiently.

The information carried in each header is necessary for a successful communication of applications over long paths comprising multiple links. This information includes the source and destination addresses, type of service, ports, protocol identifiers, etc. IP header compression is the procedure of compressing as much as possible the IP headers of a packet before transmitting it on a link. Upon reception at the other end of the link the headers are decompressed.[5] The rationale behind header compression is that most of the fields included in the header remain constant or change in a predictable manner. Therefore, it is possible either not to send them or represent them using less bits than originally required. Fig. 4 exemplifies IP header compression. As far as the IP version 4 (IPv4) is concerned, it contributes to an IP header of 20 bytes, while the UDP and RTP headers are 8 and 12 bytes, respectively. Voice packets usually carry a payload of 20 up to 60 bytes according to the voice codec used. That means that a large amount of bandwidth is used for transferring information included in the headers. However, this information hardly varies between consecutive packets, hence allowing their compression. When it comes to the IP version 6 (IPv6), a further bandwidth saving can be accomplished. IP version 6 has been designed by the Internet Engineering Task Force (IETF) to fix a number of problems that arise in IPv4, such as the limited number of IPv4 addresses. To this end, IPv6 addresses are 128 bits long, as opposed to 32 bits long IPv4 addresses, and the IPv6 header is composed of 40 octets. In this case, the headers can be compressed to as much as 3 bytes, thereby resulting in even more bandwidth savings.

So, one obvious advantage of header compression is bandwidth saving, which is very important for networks where bandwidth constitutes a resource at a premium, as is the case with wireless systems. But do there exist other benefits? One straightforward advantage is that the transmission delay is minimized, which, in turn, leads to a shorter end-to-end delay which is of significant benefit to real-time applications. When it comes to transmission over wireless lossy links, for a given bit error rate (BER) the small packet size implies that the number of received packets containing erroneous bits is reduced. It has also been shown that when FEC is complemented along with header compression, the probability of packet loss due to bit errors on wireless links significantly decreases. The study in [57] applied Reed Solomon to packet headers. That study was extended and the use of convolutional codes was proposed in order to diminish the computational complexity induced by Reed Solomon decoding [58].

A survey of header compression techniques is provided in [59]. Next, we provide a summary of the most well-known header compression techniques that have been proposed thus far. Since the focus of this paper is on real-time traffic, only those techniques that support compression of the UDP and/or RTP headers are touched upon. Moreover, we would rather not dwell upon each single detail of each compressing scheme. Instead, we accentuate the disparities among them, while providing a general description of the logic behind each technique.

7.1. IPHC

The *IP Header Compression* (IPHC) (RFC 2507 [60]) scheme was designed in 1999 with a view to working well over links with considerable packet error rates, such as wireless. IPHC provides for both TCP and UDP header compression. Specifically, the compression algorithms can be applied to IPv6 base and extension headers, IPv4 headers, TCP and UDP headers, as well as to encapsulated IPv6 and IPv4 headers. The mechanisms used for the compression of TCP headers are similar to the ones described in RFC 1144 [61]. Here, we touch only on the mechanisms related to UDP packets.

Contrary to TCP, most of the fields in the UDP header remain constant or seldom change during the session. The UDP header comprises the following fields:

- Source address (2 octets), that is, the sending port.
- Destination address (2 octets), that is, the destination port.

[5] The header compression is implemented on a hop-by-hop basis and not end-to-end.

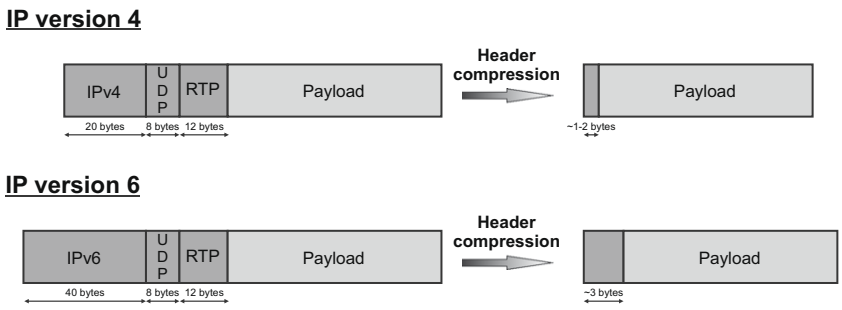

Fig. 4. Illustration of header compression.

– Length (2 octets), that is, the length in bytes of the entire datagram: header and data.
– Checksum (2 octets). This field is used for error-checking of the header and data.

The first two fields do not change, while the Length field can be inferred by the preceding subheaders, i.e., the Length field of the IP header. Therefore, the UDP header can be compressed to 2 octets since only the Checksum field should be sent.[6] Similarly, many of the fields of either the IPv4 or the IPv6 header remain constant. Therefore, there is no need to send them along with each packet. The rest of the fields can be classified into three categories: (i) the fields that carry values that can be inferred from the values of other fields. Hence these fields are not sent; (ii) those that change often but the difference from the previous values is small. So, sending only this difference requires fewer bits than sending the new value[7]; and (iii) the fields whose value changes randomly. These fields must be sent "as is" since they change unpredictably.

For UDP traffic, as soon as a new packet stream arrives, a full header carrying a *context identifier* (*CID*) is transmitted over the link to initiate compression. The CID indicates the session that the packet belongs to. Both the compressor and the decompressor store most of the fields of this full header, namely all the fields except for the ones that are included "as is". The stored fields of the full header are referred to as the *context*. Furthermore, the CID is associated with a *generation number* which is incremented by one each time the context changes. This number allows the decompressor to identify an outdated context while attempting to decompress headers. A change in one of the fields included in the context forces the compressor to send a new full header in order to update the context at the decompressor. When compressing non-TCP headers, IPHC does not use delta encoding and that makes it robust when used over lossy links since the context is not damaged when a packet is lost.

RFC 2507 also describes mechanisms to allow the decompressor to recover quickly after the loss of a packet carrying a full header. In particular, it employs a mechanism called *compression slow start* according to which full headers are also sent in the time interval between two consecutive changes in the context. The time instant at which a full header is sent relies on the number of compressed packets that have been sent after the transmission of the last full header. Whenever a change in the context occurs, this number is set to one and is doubled every time a full header is send to the decompressor. In addition to this mechanism, a full header must also be sent if the number of compressed packets exceeds a threshold. For low data rate packet stream, instead of measuring the number of compressed packets, the time interval passed since the transmission of the last full header is of interest.

In light of the above, it becomes evident that this compression scheme caters for transmission over lossy links. Nonetheless, problems may arise if a full header is received out of order. When the full header with a generation number G arrives after a compressed packet with the same generation number, then the compressed packet can be either cast away or stored temporarily until the reception of the full header. On the other hand, if a full header with a generation number G arrives before a packet with a generation number $G-1$, then the decompressor can keep both versions of the context for a while in order to be able to decompress the packet with generation number $G-1$. Last but not least, this compression scheme also provides hooks for adding header compression schemes on top of UDP, for example compression of RTP headers.

7.2. CRTP

In 1999 another header compression scheme was proposed. This scheme, however, also supports RTP headers and is described in RFC 2508 [62]. Specifically, it is tailored for the compression of the IP/UDP/RTP headers and was motivated primarily by the problem of sending audio and video over 14.4 and 28.8 kb/s dial-up modems. Moreover, it can capitalize on full-duplex links, however, it can be implemented over simplex links as well.

Like IPHR, CRTP draws heavily on the mechanisms described in [57], therefore it bears similarities to IPHR. That means that CRTP is also predicated upon the notions of context, *CID* and generation number. In addition, it makes use of a *sequence number* that is carried by both compressed and uncompressed packets and allows the decompressor to detect the loss of a packet. The sequence number

[6] If the Checksum is zero and it is likely to be so for all packets in the flow, then there is no need to send this field, thus saving two more octets in the compressed header. However, in IPv6 this field is never set to zero.

[7] This kind of compression is referred to as *Differential* or *Delta encoding*.

is incremented by one every time a packet is transmitted. Describing this compression scheme in brief, a full header is initially sent to the decompressor to establish the uncompressed header state for a particular context and allow eliding most of the fields from the compressed headers that will follow. The context can then be updated by means of a full header, a compressed UDP header or a compressed RTP header. This approach provides increased flexibility. For instance, when a change in one of the fields of the RTP header that are expected to remain constant occurs, then a full RTP header is sent to the decompressor. Nevertheless, the IP and UDP headers can still be compressed. Of course, when there is no need to update the context, the RTP header is compressed along with the IP and UDP headers to increase efficiency. It should also be noted that this compression scheme relies on the link-layer to compress the headers of the upper layers as much as possible. Particularly, the value of some fields is extrapolated from the link-layer subheaders.

As far as the RTP header is concerned, it consists of several fields that change by predictable amounts, while for some of them, the difference between two successive values remains constant. It is possible then for the decompressor to deduce the value of these fields, thus obviating the need for communicating the difference to it. However, this requires that the difference, which is called *first-order difference*, be stored in the context.[8] When the *second-order difference* of the RTP packet is not zero, that is, the first-order difference has changed, then the new values should be communicated to update the context.

Regarding error recovery, when this compression scheme is applied over simplex links, then a periodic refresh mechanism such as the one employed by IPHR should be implemented to alleviate the detrimental effect of severe packet loss. In the case of duplex links, as soon as the decompressor detects an error, it sends a special packet back to compressor indicating that the context has been invalidated. In order to avoid flooding the reverse channel, the decompressor does not send such a packet for every compressed packet that has been received in the interim.

Although the performance of CRTP is remarkable in terms of compression efficiency (headers can be compressed down to 2 bytes without UDP checksums and 4 bytes in the presence of UDP checksums), most of its advantages are nullified by the fact that it has been designed for relatively error-free links with low round-trip delay. The latter constraint is imposed by the signalling scheme that is used to expedite recovery from errors. The study in [63] proved that CRTP does not perform well over cellular radio links that are characterized by high bit errors rates. Similar observations were also made in [64,65]. The *Twice* algorithm can ameliorate the performance by making assumptions of what the lost segment was in order to repair the context without having to wait for a round-trip over the link. However, it requires the UDP checksum to be enabled, resulting in a decrease in the compression efficiency. Additionally, this compression scheme is not intended to work in conjunction with RTCP[9] (Real-time Transport Control Protocol) since the required additional complexity is undesirable.

7.3. Enhanced CRTP

Enhanced CRTP (ECRTP) was developed with the specific aim of making CRTP work well over links with high delay, high rate of packet loss and packet reordering [66]. An example of such a link is a PPP (Point-to-Point) link that is tunnelled using IP level tunnelling protocols or a link in an IP network that uses layer 2 technologies such as ATM for the access portion of the network.

The most important difference between enhanced CRTP and CRTP lies in the way that synchronization between the compressor and the decompressor is achieved. The enhanced version adopts a hybrid approach, allowing for both closed- and open-loop error recovery mechanisms. The new part that was added consists in the open-loop procedure. For enhanced CRTP to achieve robust operation the compressor sends a context update multiple times. In that way the decompressor is kept in sync. Specifically, when a change in a full or a delta (first-order difference) value occurs, the compressor includes the change in the next $N+1$ packets. Therefore, the decompressor can keep its context state in sync with the compressor using the Twice algorithm as long as no more than N consecutive packets are lost. If more than N packets are lost, then the decompressor should invalidate the context and notify the compressor via a special packet.

In addition to the open-loop mechanism, in enhanced CRTP there is also the option to send absolute values for some volatile fields rather than first-order differences. This enhancement allows synchronization between the compressor and decompressor to be maintained even when packets are lost since the context of the latter does not depend on a value carried by a previous packet. Another problem that ought to have been overcome was the inability of the decompressor to verify that context synchronization has not been lost after a packet loss when the UDP checksum is not present. To this end, a header checksum may be inserted by the compressor and removed by the decompressor after validation.[10]

As mentioned before, this compression scheme constitutes a step forward of CRTP. From this perspective, it is not a novel scheme. However, the enhancements that have been added allow it to perform much better than CRTP in lossy links with long round-trip delay, while keeping the simplicity of its predecessor. The only objection that can be raised has to do with the slightly increased overhead. Nonetheless, the enhanced performance of the scheme provides corroboration for the choice of the new mechanisms that were employed.

[8] In practice, the only fields for which it is useful to store the first-order difference are the IPv4 IP field and the RTP timestamp.

[9] Real-time Control Protocol (RTCP) is the sister protocol of RTP that is used to provide feedback on the quality of service being provided by RTP.

[10] This option is never used with IPv6 since a null UDP checksum is not allowed.

7.4. ROCCO

A highly robust and efficient header compression scheme was introduced in [67], adaptable to the characteristics of the link over which it is used and also to the properties of the packet streams it compresses. This compression scheme, which is called RObust Checksum-based header COmpression (ROCCO), is predicated on the view that the implementation simplicity of a header compression scheme is less significant than its compression ratio and robustness, and should not be a deterrent to the development of a compression algorithm.

ROCCO header compression framework does not state any exact details about how the compression is to be performed, what formats the headers should have, etc. Rather, the framework spells out general principles of how to perform header compression according to ROCCO. Specifically, for each packet the compressor computes a Cyclic Redundancy Checksum (CRC) and includes the latter in each compressed header. The reasoning behind CRC is to allow the decompressor to reliably verify if the header has been reconstructed correctly. Of course, as in all the previous algorithms, this scheme also needs to maintain a context for the compression and decompression of a packet stream. Moreover, ROCCO requires that some information such as the length of the packet can be provided by the link-layer.

It should be noted that because of ROCCO, a new working group was set up within IETF, named ROHC group. Since the time that ROHC group decided to define a new scheme, the development of ROCCO has winded up. Nonetheless, a final version was submitted on the 15th of June in 2000 to put a last and hopefully correct version of the scheme in the Internet Draft archives [67]. Albeit this version had the ROHC WG name, it should just be referred to as ROCCO 06. The outcome of ROHC group was a new compression scheme for IP/UDP/RTP headers, named Robust Header Compression (ROHC).

7.5. ROHC

RFC 3095 [68] specifies the Robust Header Compression scheme. This scheme is built upon the insights from the preceding research efforts in regard to ROCCO; however, several modifications and amendments were made as well. The incentive for this scheme arose from links with significant error rates and long round-trip times, such as cellular links. As in the previous compression algorithms, ROHC is also predicated upon the view that most of the header fields never or seldom change and shares the concept of storing common header fields for each packet stream. Only 5 fields that account for 10 octets change frequently, thus needing more sophisticated compression mechanisms. These fields are presented below:

- IPv4 Identification (IP-ID) 2 octets.
- UDP Checksum 2 octets.
- RTP Marker (M-bit) 1 bit.
- RTP Sequence Number (SN) 2 octets.
- RTP Timestamp (TS) 4 octets.

From the fields presented above, the IP-ID and TS fields are correlated with the RTP Sequence Number, which increases by one for each packet emitted by an RTP source. The M-bit barely changes; however, when it does, the new value must be explicitly communicated. As regards the UDP checksum, it is not associated with another field and, therefore, when it is enabled it must be transmitted as is. All that is needed then is to devise a function that relates the IP-ID and TS fields to the SN field. In that way only the SN field must be transmitted since decompression of the other two fields hinges only on correct decompression of the SN. Whenever the relation between SN and another field changes, additional information is sent to update the parameters of the function. It should be emphasized that ROHC makes use of *CID* and context.

According to ROHC, compression and decompression are treated as state machines which are further broken down into three states. Both state machines start with the lowest state and gradually proceed to higher states. Transition between states does not need to be synchronized between the two machines. Normally only the compressor falls back to a lower state. The decompressor resorts to a lower state only when the context has been invalidated.

For compression, the three states are the *Initialization and refresh* (*IR*), *First-Order (FI*) and *Second-Order (SO*). The compressor will always operate at the highest possible compression provided that the decompressor has all the information required to handle a header compressed according to that state. In the IR state, full packet headers are sent in order to initialize the context at the decompressor's side. In the FO state, the compressor rarely sends information about all dynamic fields, and what is more, the information sent is usually compressed at least partially. Concerning static fields, only a few of them can be updated. Consequently, the compressor enters this state whenever the headers of the packet stream do not conform to their previous pattern. Finally, the compressor enters the SO when it is confident that all the information about irregularities in the packet stream has been communicated to the decompressor. This is the state where header compression is optimal in the sense that, aside from the fields that cannot be predicted, in this state the decompressor is able to determine all the header fields via extrapolation from the SN's value. Decisions about transitions between the various compression states are taken by the compressor on the basis of variations in packet headers, feedback from the decompressor and periodic timeouts.

As far as the states of the decompressor are concerned, these are the *No Context*, *Static Context* and *Full Context*. Initially, the decompressor is in the *No Context* state since it has no information about the characteristics of the packet stream. As soon as the context is set up and a packet has been decompressed correctly, it can transit directly to the *Full Context* state and only upon repeated failures will it transit back to lower states. When that happens, it first falls back to the *Static Context* state. The correct decompression of a packet sent at the FO state is enough to send the decompressor to the *Full Context* state again. The decompressor will resort to the *No Context* state only if

the decompression of several packets sent in the FO state fails. It is evident, that the maximum compression efficiency can be accomplished only when both the compressor and decompressor are in the highest states. In this occasion, the header will be compressed in most cases to as low as 1 byte if UDP checksum is disabled.

Except for the aforementioned states of the compressor and the decompressor, the ROHC scheme has also three modes of operation, called *Unidirectional*, *Bidirectional Optimistic* and *Bidirectional Reliable*. The *Unidirectional* mode of operation is suitable for simplex links. According to this mode, the transitions between the compressor states are governed by periodic timeouts and irregularities in the way some header fields change. The *Unidirectional* mode results in lower compression ratios with regard to the *Bidirectional* modes on account of the lack of any kind of feedback. However, the ROHC scheme rules that compression must always start in this mode. Transition to any of the *Bidirectional* modes can be performed as soon as a packet has reached the decompressor and the latter has replied with a feedback packet indicating that a mode transition is desired.

In the *Bidirectional* Optimistic mode a feedback channel is used to enable the decompresssor to send error recovery requests and acknowledgments of significant context updates back to the compressor. In this mode, periodic refreshes are not employed. The aim of this mode is to increase compression efficiency. The price to pay is, however, a higher context invalidation rate, especially when long error bursts occur. The third mode prevents loss of context synchronization if BER is not really high by resorting to a more intensive usage of the reverse channel, as well as to a stricter logic at both the compressor and decompressor. In this mode all context updates are acknowledged.

As a whole, ROHC is a powerful scheme that performs well over lossy links with long round-trip delay, while achieving high compression ratios at the same time. Apart from the mechanisms described in the previous paragraphs, the increased compression efficiency can be ascribed to its ability to use different encoding schemes for each header field. The authors of [69,70] assessed the performance of ROHC in wireless links using an evaluation methodology based on objective voice quality metrics, namely metrics that can be mathematically calculated. It was found that this compression scheme ameliorates voice quality especially for high bit error probabilities. RFC 3242 [71] defined additional mechanisms that capitalize on functionality provided by the link-layer in order to increase compression efficiency by completely eliminating the header for most packets during optimal operation in the *Unidirectional* and *Bidirectional Optimistic* modes. These mechanisms were extended in RFC 3408 [72] in order that zero-byte compression can be supported in *Bidirectional Reliable* mode as well. Moreover, it should be noted that the implementation of the compressor has several ways at hand to handle the RTCP stream, something that is possible neither with CRTP nor enhanced CRTP. However, this versatility comes at the expense of an increased complexity that cannot be overlooked.

Table 4
Comparison of the header compression schemes.

	IPHC	CRTP	ECRTP	ROHC
Maximum compression	2 bytes	2 bytes	2 bytes	1 byte
Robustness to errors	Low	Low	High	High
Robustness to long delays	Low	Low	High	High
Compression ratios	High	High	Medium	High
Robustness to reordering	No	No	Yes	No
Complexity	Low	Low	Low	High

7.6. Comparison of header compression algorithms

Table 4 provides a summary of the salient features of the aforementioned compression schemes. For the sake of completeness, we also include the IPHC scheme in this table although it is not tailored for RTP header compression. Moreover, ROCCO was not considered since it can be viewed as the ROHC's predecessor. A first comparison of header compression schemes was carried out in [73,74]. In that study CRTP was compared to ROCCO, the precursor of ROHC. The latter scheme proved much more efficacious in terms of compression rates and robustness against the loss of consecutive packets. However, both schemes were superseded by ECRTP and ROHC, respectively, whose design was based on the lessons learned from previous header compression algorithms. In [64], the ECRTP and ROHC compression schemes were evaluated under the consideration of a CDMA2000 air interface. Simulation results revealed that ROHC attains higher compression rates under any mode. In addition, the packet error rate (PER) of both schemes turned out to be similar with the exception of the *Unidirectional* mode of ROHC where PER was much worse due to the lack of feedback to the compressor. Therefore, we reach the conclusion that the most appealing compression schemes appear to be ECRTP and ROHC.

In conclusion, the choice of the compression scheme is mostly dictated by the type of the channel over which the scheme should work. IPHC and CRTP are apt for channels with low delay and low round-trip time. However, they might not work well when packets arrive out-of-order, a case that is strongly dependent upon the underlying routing scheme. On the other hand, ECRTP and ROHC are suitable for lossy links with high round-trip delay. Nonetheless, the price to pay in the case of ECRTP is the decrease in compression gain, whereas in the case of ROHC is its high complexity.

8. Signaling protocols

For the establishment of telephone calls over the Internet, a signaling protocol is of paramount importance because it enables network components to communicate with each other, set up and tear down calls. In the case of IP telephony, a call can be defined as the multimedia session between two or more participants, while signaling associated with a call is referred to as a connection. The role of a signaling protocol can be broken down into four functions:

- *User location:* The caller first has to find the location of the callee.
- *Session establishment:* The called party decides on whether to accept, reject or redirect the call.
- *Session negotiation:* The endpoints involved in the call should agree upon a set of properties for the session.
- *Call participant management:* It allows endpoints to join or leave an existing session.

Signaling protocols for VoIP have been the focal point of debates since the very first days of VoIP industry. Both ITU-T and IETF (Internet Engineering Task Force) embarked on the development of VoIP signaling protocols in the last years of the previous decade. H.323 and SIP (session initiation protocol) represent two basic protocol architectures in the area of multimedia over IP. The former was developed by ITU-T, whereas the latter is the fruit of the efforts made in IETF. Albeit when discussing these protocol suits, VoIP immediately springs to mind, these architectures provide far more services than just setting up voice calls. Besides these two well-known protocols, the last years three new protocols have come to light; the IAX protocol, the Media Gateway Control Protocol (MGCP) and the MeGaCo/H.248 protocol. In addition to these protocols, we also discuss the protocol that Skype, a VoIP service, uses. Next, we describe a general architecture for the integration of PSTN and VoIP network and provide a concise description of these protocols with the aim of casting light to the main characteristics of each one.

8.1. Interworking with circuit-switched telephony networks

The majority of voice calls are still carried over circuit-switched networks, namely PSTN. Thereby, one important issue that arises is interworking between VoIP networks and circuit-switched networks. Aside from voice data, signaling data is also exchanged between these two types of networks. As far as signaling in circuit-switched networks is concerned, the Common Channel Signaling System No. 7 (SS7) is used, which has been defined by ITU Telecommunication Standardization Sector (ITU-T). Thus, direct connection between PSTN and Voice networks is not feasible. SS7 is the means through which elements in telephony networks exchange information. SS7 messages are carried out-of-band on dedicated channels for two reasons. The first one is for achieving faster call setup times, since there is no need for multifrequency signaling tones which are used in in-band signaling, while the second one if for avoiding fraudulent network usage, since users do not have access to these channels. The primary tasks of SS7 are call setup, management and tear-down. However, it also provides for call features such as call forwarding and calling number display. In an SS7 network there exist three kinds of endpoints: Service Switching Points (SSPs), Signal Transfer Points (STPs) and Service Control Points (SCPs). SSPs are switches that connect voice circuits. An SSP sends signaling messages to other SSPs to setup, manage and release voice circuits required to complete a call. STPs route network traffic to an outgoing signaling link based on routing information contained in the SS7 message, while SCPs provide SSPs with access to databases for additional routing information used in call processing. All endpoints in an SS7 network are connected by signaling links. These links are classified into six categories according to the type of endpoints that they connect.

So, is interworking between VoIP networks and PSTN feasible? One would expect that interworking with an old telephony infrastructure would be a daunting challenge. However, the widespread adoption of SS7 constitutes a stepping stone to the successful interworking of VoIP systems with PSTN. Fig. 5a illustrates one potential way a service provider can use an IP-based network for telephony services. A VoIP Gateway is a network device which transfers voice and fax calls, in real-time, between an IP network and PSTN. It can be broken down to the media gateway and the signalling gateway. The former compresses and packetizes voice data and delivers compressed voice packets to the IP network at one end of the endpoint and does the exact opposite procedure at the other endpoint. The latter provides transparent interworking of signalling between circuit-switched and IP networks. The signaling gateway may terminate SS7 signaling or translate and relay messages over an IP network to a media gateway controller or another signaling gateway. SoftSwitch is a software-based switching platform, as opposed to traditional hardware-based switching center technology, which is used to bridge a traditional PSTN and VoIP by linking PSTN to IP networks and managing traffic that contains a mixture of voice, fax, data and video. It is also known under different terms such as media gateway controller, call agent and gatekeeper. SoftSwitches are able to process the signaling for all types of packet protocols. Signaling information is send to the media gateway controller by the Gateway. Describing in brief the call setup message flow, at call setup time the PSTN switch seizes a trunk to the Gateway and signals the near-end SoftSwitch, through the Gateway, with the call information. The SoftSwitch in turn determines the zone the called number is in and sends a setup message to the SoftSwitch that is responsible for that zone including the IP address of the originating Gateway. Upon the reception of this message, the terminating SoftSwitch is shouldered the process of identifying the terminating Gateway and signaling the far-end PSTN switch with a trunk ID from the Gateway. Then, the PSTN switch should complete the call to the number dialed. Another variant of the aforementioned architecture is depicted in Fig. 5b. In this figure the terminating terminal is an IP phone directly attached to the packet network. The session manager is involved in the setup/tear-down signaling through a protocol such as H.323 or SIP. In the next subsections, each signalling protocol is also reviewed with regard to call features such as call forwarding and call transfer, which are deemed necessary for smooth interworking with PSTN, as well as the architecture that it employs for the control of these services.

8.2. The H.323 protocol suite

The first version of H.323 was ratified in 1996. Since then several amendments have been proposed and embraced [75]. In fact, H.323 is not a single protocol. Rather

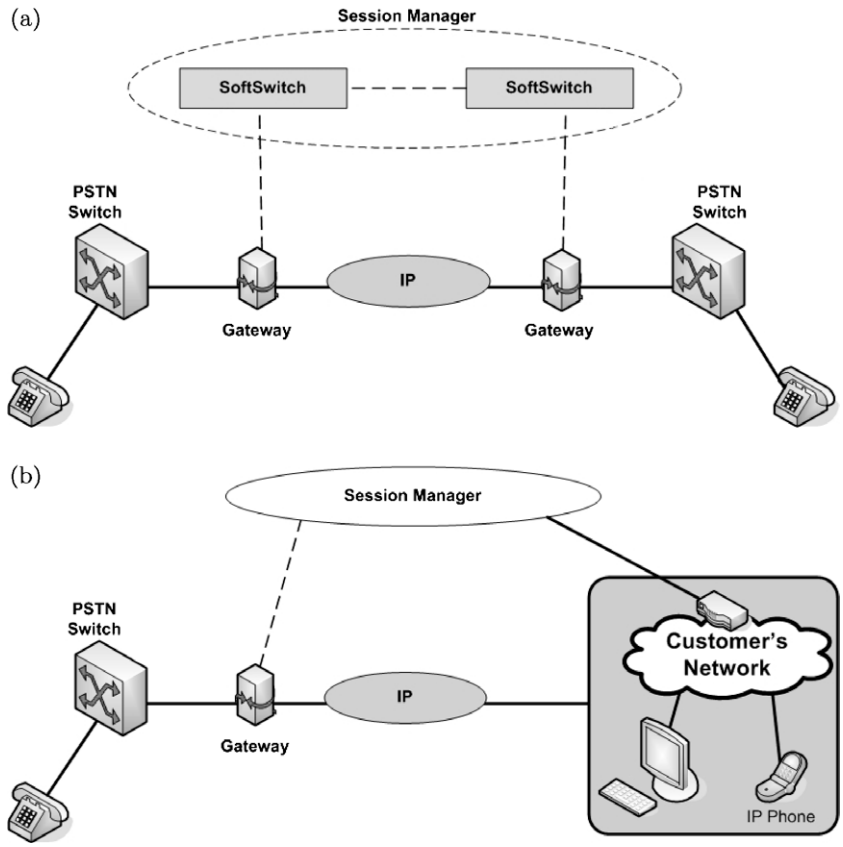

Fig. 5. Generic VoIP architectural solutions.

it is a suite of different protocols that are used to support audio, video and data applications. H.323 consists of three types of interactions. The first one refers to the administrative operations of a user and is carried out using the Registration Admission and Status protocol (RAS). The second one is responsible for signaling regarding call setup and tear-down. To this end the H.225 protocol is employed. The third type of interaction addresses terminal's capabilities negotiation and call control and is covered in H.245.

Features in H.323 can be categorized into three classes: local features, network-based features and supplementary services [76]. Local issues include call history, address book as well as other features that can be implemented at the endpoint, thereby obviating the need for specific signaling. Network-based features comprise authorization, call admission, etc., that is, features that require centralized control. Last but not least, the third class consists of a supplementary set of services that necessitate special signaling. Supplementary services are user-perceived features that enhance the call with specific functionality, such as call forwarding, call transfer and call completion, which is required for interworking with the PSTN.

8.2.1. H.323 network architecture

In this subsection we describe the main elements of a network architecture that relies on the H.323 umbrella protocol, namely terminals, Gateways, Gatekeepers and Multipoint Control Units (MCUs). A general network architecture is depicted in Fig. 6.

A terminal (also referred to as endpoint) is simply the client endpoint and must support the following four protocols: H.245, Q.931, RAS and RTP. RAS is the protocol used by terminals and gateways for registration and admission purposes. As mentioned earlier, a gateway is the network device that provides two-way communications, in real-time, between an IP network and legacy telephone networks (i.e., PSTN). In other words, it is the interface between the PSTN and the Internet and performs the translation between different transmission formats. The basic functionalities of RAS are to help a terminal find a gateway and register with it, request permission from a gateway before the beginning of a call, and disengage a terminal from the gateway. On the other hand, H.225, which adopts a subset of Q.931[11] messages and parameters, performs all signaling that is required to set up and maintain connection between two terminals. Unlike RAS which uses UDP, Q.931 is transmitted over TCP. As regards H.245, it is used to convey end-to-end control signaling between the communicating terminals. These control messages carry information regarding the capabilities of each endpoint and defines procedures for mapping logical chan-

[11] Q.931 is the ITU-T recommendation for call signaling in ISDN.

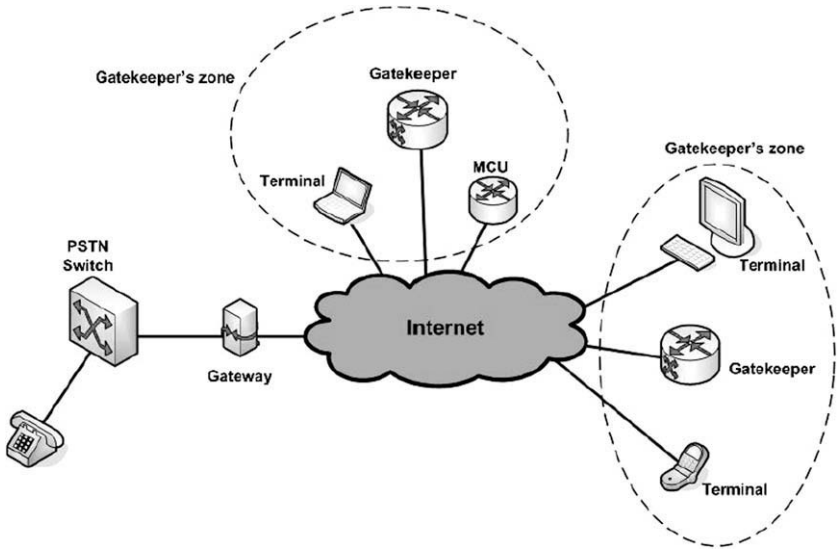

Fig. 6. The H.323 network architecture.

nels. Moreover, another function of H.245 is to identify which terminal will act as a master and which one as a slave. This is necessary in order to prevent conflicts that may arise when both terminals initiate similar events simultaneously.

The third component in the H.323 architecture is the gatekeeper. The gatekeeper performs the call control functions and administration of policy for terminals within its zone. All terminals must be registered with one gatekeeper. The mandatory functionalities that gatekeepers provide are listed below:

- *Address translation:* The translation of an alias address to a transport address.
- *Admission control:* Access to the network can either be granted or denied based on call authorization, source and destination addresses, etc. This service makes use of RAS messages and does not rule how network resources should be utilized.
- *Bandwidth control and management:* Control of the number of terminals that can exist in the network at the same time.
- *Zone management:* The terminals, gateways and MCUs managed by a single gatekeeper.

Aside from the compulsory services, a gatekeeper can also provide some optional functions:

- *Call authorization:* The gatekeeper can accept or reject a call based on several factors such as the time of day and restricted access to particular terminals or gateways.
- *Call control signaling:* The gatekeeper can decide on whether it will process all call signaling associated with the terminals within its zone or allow endpoints to exchange directly signaling messages among them.
- *Call management:* The gatekeeper keeps information about the status of ongoing calls such as the bandwidth used. Call redirection is also included in the tasks assigned to this service.

The last component of the H.323 network architecture is MCU, which provides conference support for three or more terminals. It comprises one multipoint controller (MC) and optionally one or more multipoint processors (MP). The former entity performs conference control by controlling what media streams go where, whereas the latter entity mixes, switches and processes media streams.

8.2.2. Supplementary services

The H.323 service architecture provides three main models for supplementary service control: distributed feature control (H.450), stimulus feature control (H.323 Annex L), and application-layer feature control (H.323 Annex K) [76]. The distributed feature control uses a decentralized function, thus supplementary services make use of the H.450 signaling without involving centralized network control. At the other extreme, the stimulus feature control relies on a centralized approach. As far as the application-layer feature control is concerned, it allows the development and deployment of new services without updating the H.323 protocol and endpoints. To this end, a service control session can be established after exchanging the relevant information (e.g., session ID, URL for service control, etc.) in RAS or H.225-CC messages.

8.2.3. Call stages

Albeit the process of establishing and maintaining a call in a H.323 network architecture is rather complex, it can be broken down to five stages [77]:

- Discovery and registration
- Call setup

- Call-signaling flows.
- Media stream and media control flows.
- Call termination.

During the discovery and registration stage, the gatekeeper that the caller is registered with commences a procedure to determine the gatekeeper with which the terminal must communicate. Once all the procedures of this stage are completed, the call setup stage follows where the gateways communicate directly to establish the connection. During the next stage, the gateways exchange messages regarding their capabilities and a media channel is opened in each direction that will be used for transferring the media stream toward the other gateway. In the media stream and media control flows stage, the RTP encapsulated media stream is sent over the links established in the previous stage. Along with these packet streams, control information about the RTP flows are also exchanged by means of RTCP. During this stage endpoints may seek changes in the amount of bandwidth initially requested and allocated to the corresponding media stream. Last but not least, when call termination is requested by either a gateway or a terminal, the media stream is stopped and the media channels are closed. Moreover, both gateways disengage with the gatekeeper via RAS. For more information on these stages, the avid reader is referred to [77] where an in-depth description of each stage is provided.

8.3. SIP

SIP is an application-layer protocol that was initially specified by the IETF Multiparty Multimedia Session Control Working Group (MMUSIC WG) in 1999 and updated by the SIP WG in 2002. SIP, which is delineated in RFC 3261 [78], is used for creating, modifying and terminating sessions with one or more participants, and was designed to be independent of the underlying transport protocol. As in H.323, features in SIP are also classified into three similar categories, namely local features, network-based features such as authorization and supplementary services. The main functions of this signaling protocol are: (i) location of resources/parties; (ii) invitation to service sessions; and (iii) negotiation of service parameters. For conveying information about the media content of the session SIP relies on the Session Description Protocol (SDP).

SIP is similar to HTTP (HyperText Transfer Protocol) and shares some of its design principles. In particular, it adopts a client/server (request/response) architecture in which requests are generated by the client and sent to the server. The server then processes the requests and sends a response to the client. Like HTTP, SIP is based on text-based messages which are either requests or responses exchanged between the clients and the servers. The most important types of requests are the INVITE request that is used to invite a user to a call, the ACK that sends the caller to the callee to simply acknowledge the receipt of the latter's response, and the BYE request that is used to terminate the connection between two users in a session. In addition to these types, three other kinds of requests can be identified, that is, the CANCEL, the OPTIONS and the REGISTER requests. The CANCEL request is used to countermand any pending searching for a user; however, it does not tear down an ongoing call. The OPTIONS request just queries the capabilities of servers. Finally, the REGISTER request is employed to register a user with a SIP server.

8.3.1. SIP network architecture

A simple paradigm of a SIP architecture is illustrated in Fig. 7. The main components of SIP are the user agents and networks servers. User agents are the SIP's endpoints that make and receive calls. A user agent can function in two modes: either as a user agent client (UAC) that initiates SIP requests or as a user agent server (UAS) that receives requests and responds on behalf of the user. In practice, a SIP endpoint (for instance, an IP phone) can act as both a UAC and a UAS. However, it functions only as one or the other per transaction depending on whether or not it initiated the request.

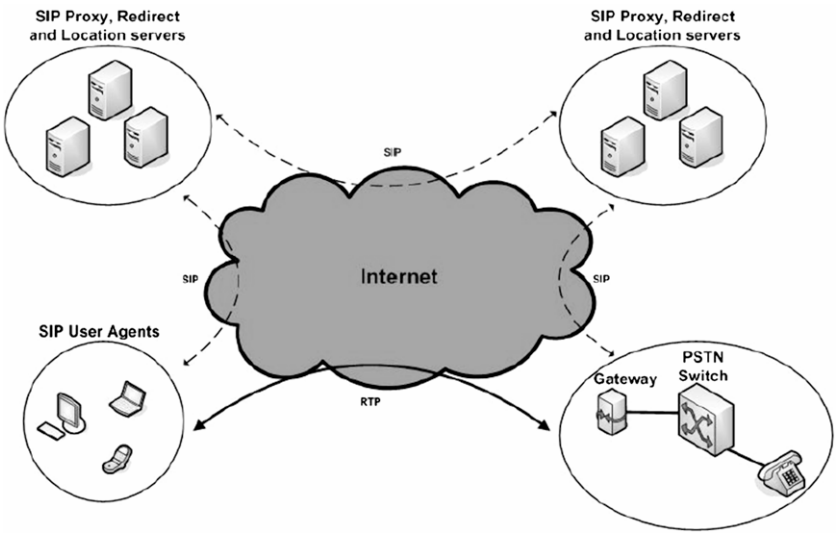

Fig. 7. SIP network architecture.

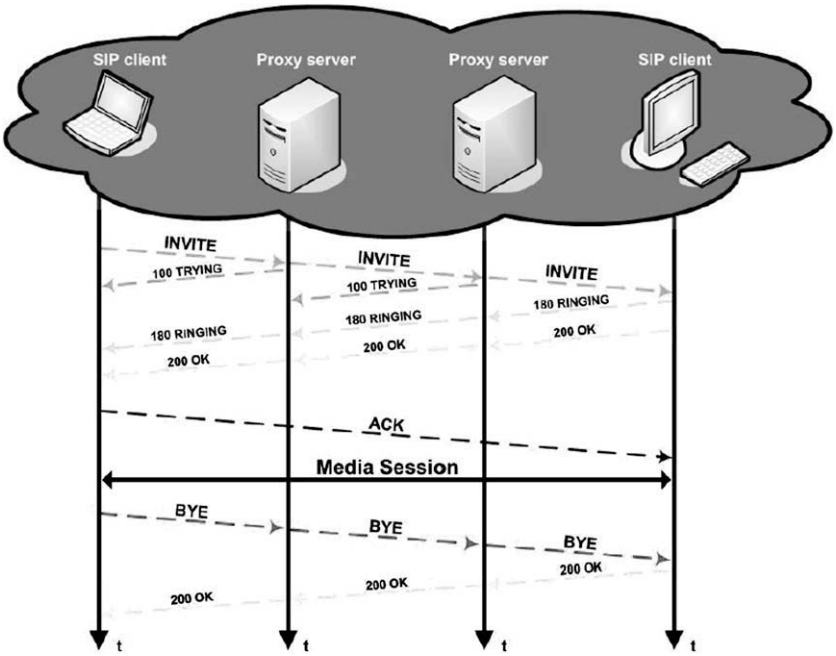

Fig. 8. Call setup and tear-down in a SIP architecture.

As far as network servers are concerned, there exist four different types of them in a network: proxy servers, redirect servers, location servers and registrar servers. A proxy server receives the requests generated by user agents and decides to which server a request should be forwarded. A request will usually traverse many servers before reaching its destination. The purpose of a redirect server is different from that of a proxy server. A redirect server does not forward requests. Rather it notifies the calling party of the location of the callee. To do so, it contacts a location server that keeps information about the called party's possible location. As for the purpose of registrar servers, they accept REGISTER requests from user agents and are usually co-located with either proxy servers or redirect servers. Finally, as in the case of the H.323 architecture, a gateway can also be employed to bridge SIP endpoints with other types of terminals.

8.3.2. Supplementary services

Although supplementary services were initially left aside, steps have been taken toward the provision of a set of mechanisms for a plethora of applications, encompassing many features, including supplementary services. Nevertheless, some of these new features require protocol extensions. It should also be emphasized that on account of the fact that SIP relies on intelligent terminals, stimulus control is considered to be outside the scope of SIP due to its centralized approach [76].

8.3.3. Call stages

Fig. 8 aims to exemplify how a call is established and torn down in a SIP architecture. The calling party first sends an INVITE request toward the called party. This request traverses some proxy servers before reaching its destination. Each proxy server that receives this request sends back an 180 TRYING response message, indicating that the request is being processed. When the UAS of the called party receives the request it starts "ringing" and thereupon sends an 180 RINGING response to the UAC of the calling party. The proxy servers that receive the 180 RINGING response forward it back to the caller. When the call is accepted by the callee a 200 OK response is sent back to the calling party. The calling user agent then replies to that message with an ACK which is sent directly to the callee. Then the two parties can commence communication. Had the callee decided to reject the call, a 603 DECLINE message would have been sent to the caller (when busy, a 600 BUSY response is forwarded). Supposing that the time that the caller wants to hang up comes, a BYE request is sent via the intermediate proxy servers to the callee, and the latter replies with a 200 OK response to indicate that the request accomplished its purpose.

8.4. IAX

IAX stands for Inter-Asterisk eXchange protocol and represents an application-layer protocol originally developed by Mark Spenser to enable VoIP connections between Asterisk[12] servers or in general between servers and clients that use it. The original IAX protocol has been deprecated virtually universally in favor of the IAX version 2, which is referred to as IAX2 [79]. Aside from Asterisk servers, this protocol is also used by FreeSWITCH servers.[13] IAX2 is a robust yet lightweight protocol. While SIP and H.323 were

[12] Asterisk is an open-source implementation of a telephone private branch exchange (PBX).

[13] FreeSWITCH is an open-source telephony application written in the C programming language.

developed to be functional for a multitude of multimedia applications, IAX2 was highly optimized for VoIP calls where low overhead and low bandwidth consumptions are the main priorities. However, it is general enough to handle most common types of media streams.

The basic design principle upon which this protocol is built is the use of the same UDP port (usually port 4569) for any type of IAX2 traffic. In other words, it multiplexes signaling and media streams over a single UDP stream between two hosts. By doing so, it works better than the other commonly deployed protocols behind network address translation[14] (NAT) firewalls since it does not use different network addresses for signaling and media messages. Moreover, according to IAX2 the end-to-end path can be divided into several segments and a different signaling protocol can be implemented in each one of them. Further, IAX2 allows multiple media streams between the same pair of peers to be multiplexed into a single trunk call coalescing media payload into a combined packet, thus diminishing overhead without creating additional latency.

IAX2 also supports call encryption on a call-by-call basis using the Advanced Encryption Standard (AES) [80]. Despite the advantages of IAX2, it enjoys little penetration in the market since vendors are usually disinclined to invest in products that are not neatly defined, since it is only recently that IAX2 has been submitted to IETF as an Internet draft, while during the previous years the source code was considered to be the documentation for this protocol.

8.4.1. IAX2 network architecture

IAX was initially developed to enable VoIP in the Asterisk architecture. In particular, IAX allows calls to be switched between Asterisk systems or devices (like IAX telephones). Hereinafter, an Asterisk server or device that implements the IAX2 protocol is referred to as the peer. Hence, peers are the main elements of an IAX2-based architecture and communicate with each other via IAX2 messages that are carried within frames. Frames can be classified into three main categories:

- Full frames that carry signaling or media data in a reliable manner, requiring immediate acknowledgement upon receipt. In general, they are used to control initiation, setup, and termination of an IAX2 call, but they can also be employed to carry stream data, though this is generally not optimal.
- Mini frames whose sole purpose is to carry media stream data on an already-established IAX2 call. They are transmitted in an unreliable manner since voice data are typically sent in real-time, therefore it does not make sense to retransmit lost frames.
- Meta frames used for call trunking or video stream transmission in an unreliable manner.

In order for an IAX2 peer to be reachable by other peers, the calling peer needs the network address of the called peer. This address can be manually determined through a shared directory or configured using the IAX2 protocol which provides a facility for one peer to register its address and credentials with another (which is termed Registrar) so that other peers can locate it. However, this facility is optional. In addition, peers usually maintain their own dialplan, that is, a set of rules for associating provided names and numbers with a particular called party. Nonetheless, IAX2 also provides for users that may want to dial from a peer that does not switch its own calls.

8.4.2. Call stages

Fig. 9 depicts the exchange of messages between two IAX2 peers. Full frames are represented by solid lines, whereas dash lines denote mini frames. A call is initiated with a NEW message indicating the destination "number" or name of the called party in the remote peer. Upon receipt of this message, the remote peer can reply with either a credentials challenge (AUTHREQ message) if required, a REJECT message, or an ACCEPT message. The AUTHREQ message indicates the permitted authentication schemes and results in the sending of an AUTHREP message with the requested credentials. The REJECT message indicates that the call cannot be established, while the ACCEPT message indicates that the path between the two peers is set up.

Assuming that the authentication succeeds, an ACCEPT message is sent to the originating peer (otherwise a REJECT response is sent), notifying it of the desired codec. This peer has to reply with an ACK and waits for one of the possible call control messages. These messages include RINGING, ANSWER, BUSY and PROCEEDING. Typically the first call control message is RINGING which indicates that the called party is being alerted to the call. The PROCEEDING message should be sent to a calling party when their call request has not reached the called party and is being processed by a further network element. Supposing that at some time the called party accepts the call, an ANSWER message is sent to the originating peer. Recall that all these messages are sent within full frames, thereby requiring acknowledgement. Upon receipt of the ANSWER message, the communication channel must be opened in both directions before the transmission of voice data commences.

As stated earlier, voice data is carried within mini frames. Mini frames carry the low 16 bits of the peer's 32-bit long timestamp. Therefore, full frames should be exchanged periodically to synchronize the 32-bit long timestamp when the 16-bit long timestamp overflows. This approach facilitates both efficiency and reliability. When one of the two communicating parties wants to tear down the call, it sends a HANGUP message to the other peer. Upon receipt of a HANGUP message, an IAX2 peer must immediately respond with an ACK and then tear down the path at its end.

8.4.3. Supplementary services

Supplementary services are not really touched upon in the Internet draft submitted to IETF. Nevertheless, it is almost certain that as the popularity of IAX2 grows, support for supplementary services will be provided to enable smooth interworking with the PSTN.

[14] Network address translation is called the process or rewriting the source and/or destination addresses of IP packets as they pass through a router or a firewall. By doing so, it becomes possible for multiple hosts on a private network to access the Internet using a single IP address.

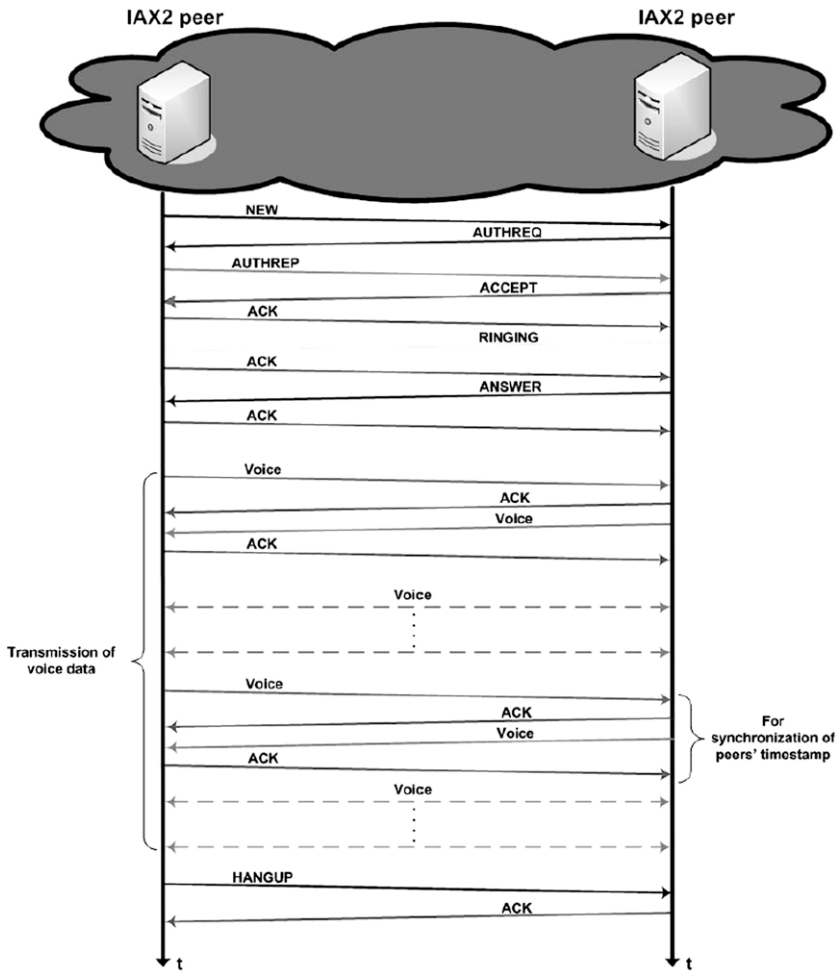

Fig. 9. Complete end-to-end IAX2 media exchange.

8.5. MGCP

MGCP is fully defined in RFC 3435 [81] and can be viewed as the merger between the Cisco Systems and Bellcore's Simple Gateway Control Protocol (SGCP[15]) and the level 3 Technical Advisory Committee's Internet Protocol Device Control (IPDC[16]) specification. Like these two protocols, MGCP was geared towards a simplified design and centralized call control, hence being suited for a large-scale IP telephony deployment.

8.5.1. MGCP network architecture

The main entities in a MGCP-based architecture is the call agent, also referred to as Media Gateway Controller (MGCs), the media gateway, whose role is to provide conversion between audio signals carried on telephone circuits and data packet carried over the Internet, and a signaling gateway, if the network is connected to an SS7 controlled network. A signaling gateway is solely responsible for translating signaling messages between one medium (PSTN) and another (IP). Contrary to SIP and IAX2, which are peer-to-peer protocols, MGCP is a centralized protocol. The call agent is in charge of establishing, modifying and terminating calls, namely, it signals to, and controls, media gateways in order to connect and control VoIP calls. Therefore, it presents itself as an H.323 gatekeeper or an H.323 endpoint to the H.323 systems.

MGCP works on a master/slave basis in that media gateways execute commands that are sent to them by call agents. It is organized as a set of transactions, each one of which is comprised of a command and a response (i.e., an acknowledgement). An advantage of this architecture is that new services can be introduced without having to swap or upgrade media gateways.

MGCP assumes a connection model within the media gateway according to which a connection has two entities;

[15] SGCP was a protocol published in 1998 as a part of the "Call Agent Architecture" at Telcordia. In that architecture a call agent controls media gateways and receives telephony signalling requests through a signalling gateway.

[16] IPDC was designed to provide a medium to bridge VoIP networks and traditional telephony networks.

endpoints and connections. Endpoints are the physical, or at times, logical interfaces that either initiate or terminate a VoIP connection. A connection is the association between two endpoints, which may reside in the same or different media gateways, with the purpose of transferring data between these endpoints. There exist two flavors of connection: point-to-point and point-to-multipoint. The call agent is only responsible for the management of connections in a MGCP architecture.

8.5.2. Call stages

To set up and tear down connections between endpoints that are located on media gateways that are managed by the same call agent, the following three steps must be taken:

- The call agent first asks the first gateway to create a connection on the first endpoint. The gateway then allocates resources to that connection and sends a response back to the call agent including a session description that contains pertinent information, such as IP address, UDP port and codec parameters, required by third parties to send packets toward the newly created connection.
- The call agent then sends the session description of the first gateway to the second gateway and also asks it to create a connection on the second endpoint.
- Upon receipt of the response from the second gateway, the call agent uses a "modify connection" command to provide the first endpoint with this second session description. Once this is carried out, communication can proceed in both directions. The call agent can change the parameters of the connection at any time by sending a "modify connection" command.

To tear down a connection, the call agent sends a "delete connection" command to the gateways.

When the endpoints are located on gateways that are managed by different call agents, these two call agents shall exchange information by means of a call agent to call agent signaling protocol in order to be synchronized and send coherent commands to the gateways under their control. MGCP does not specify such a signaling protocol for synchronizing call agents. However, synchronization between call agents can be attained through an existing protocol such as SIP or H.323. From this perspective, MGCP can be considered as a complementary rather than a competitive protocol to these protocols.

8.6. MeGaCo/H.248

MeGaCo/H.248 is often referred to as the successor to MGCP and represents a common cooperative effort between IETF and ITU to develop a protocol for the control of elements in a physically decomposed media gateway, which enables separation of call control from media conversion. In IETF the name MeGaCo is used [82], which is a contraction of "Media Gateway Controller", while in ITU the same protocol is known as H.248[17] [83]. This was the first time that these two groups have worked on a standard together and is indicative of the future meshing of telecommunications and the Internet. Hereinafter, we refer to this protocol as H.248.

H.248 was built on the foundations of MGCP. Therefore, many concepts proposed in MGCP have found their way into the H.248 standard, though the latter offers several enhancements and features that MGCP does not. In this section we aim to accentuate the new functionalities that H.248 brought with its arrival.

Instead of endpoints and connections, the connection model in H.248 is based on terminations and contexts. A termination sources and/or sinks one or more streams. In a multimedia conference, a termination may source or sink multiple media streams. The media stream parameters are encapsulated within the termination. A context is essentially a grouping of the terminations associated with a call. Idles lines are represented by termination in a special type of context called the null context. In comparison with the MGCP connection model, the H.248 connection model considerably simplifies call setup. For instance, setting up a multi-party conference using H.248 merely involves adding several terminations to a context, whilst in the case of MGCP several connections must be established to a special type of endpoint. Even though the connection model differs in H.248 compared to MGCP, the semantics of the commands in these two specifications bear significant similarity. Specifically, there is one-to-one mapping between the commands of the two protocols. Last but not least, H.248 allows a single termination to have multiple media streams that may be transmitted or received on different bearer channels, thereby enabling support for H.320 multimedia terminals.

8.7. Skype

Skype arrived in 2003 from the people that developed the KaZaa file sharing system. Were we to weight protocols by the number of people that use them, then Skype would top the list since it boasts more than 6 million online users at any time. Albeit the signaling protocol used to connect Skype users is proprietary to Skype, it turned out to be pretty efficacious since it deals successfully with NAT and firewall traversal. Unfortunately, it is rather hard to procure details about it, and therefore, it cannot be compared to the aforementioned protocols on a fair basis. However, it is worth devoting a few lines to how Skype works.

8.7.1. Skype network architecture

In Skype, there exist two types of nodes, ordinary hosts and super nodes. An ordinary host is a Skype application that runs on a user's computer, while a super node is an ordinary host's endpoint on the Skype network and should have a public IP address. An ordinary host must be connected to a super node and must register itself with the Skype login server. Aside from this central server, all the

[17] H.248 was renumbered when revised on March of 2003. The main body of H.248, Annexes A–E, and Appendix I were included in H.248.1, while subsequent Annexes were sequentially numbered in the series.

other entities are distributed and online or offline information is stored and propagated in a decentralized fashion. Skype employs TCP for signaling, while media traffic is transported either via UDP or TCP. Nevertheless, signaling and media traffic are not sent on the same ports.

8.7.2. Call stages

The most critical process of the Skype's operation is Login. During this process, a Skype client authenticates its user name and password with the login server, advertises it presence to other peers and determines the type of NAT and firewall it is behind. It takes place when a user starts the application. Concerning call setup and tear down, three different scenarios can be identified [84]. When both users are on public IP addresses, the caller establishes a TCP connection with the callee and signaling messages are exchanged over TCP. If one of the users is behind port-restricted NAT, then signaling messages do not flow directly between the caller and the callee. Instead, communication is achieved through an intermediary super node that receives messages from the caller over TCP and forwards them to the callee again over TCP. In both these scenarios, voice packets are transported over UDP. When both users are behind port-restricted NAT and UDP-restricted firewall, an intermediate super node is used to receive and forward messages between the communicating clients over TCP. However, in this case, voice packets are also sent over TCP. For call tear down, signaling information is also exchanged over TCP. However, in the last two scenarios the messages pass through a super node.

8.8. Comparison of signaling protocols

Signaling has been one of the key areas of VoIP technology since its inception, thus the slew of protocols that have been proposed for VoIP as yet. H.323 came out just at the right time for the fledgling VoIP industry. Therefore, it is no surprise that the VoIP market has been inundated with products that support H.323. Usually H.323 is found in organizations that have employed VoIP technologies and solutions for some period of time, thus resorting to H.323 since it is the oldest of the protocols we have discussed thus far. On the other hand, SIP is younger than H.323 and gained significant momentum since its infancy due to its lower complexity. These two protocols have been dominant in the VoIP technology arena, therefore most of the studies in this field focused on the differences between these two protocols.

The first comparison of SIP and H.323 for internet telephony was made in [85]. The protocols were compared in terms of complexity, extensibility, scalability and features. SIP was found to provide a similar set of services, while being less complex. In addition, the authors of that study drew the conclusion that SIP provides richer extensibility and better scalability. In [86] the effect of different queuing policies on the performance of SIP and H.323 under extreme traffic congestion was investigated. SIP fared better than H.323 in that it attained a higher percentage of successful session establishments. A further comparison between SIP and H.323 can be found in [76], where the protocols are compared according to their standardization philosophy, the supported and supplementary services, interoperability of services and features, and interworking with PSTN. H.323 standards are more specified in the sense that they include solutions for cross-sectional factors that are not touch upon in SIP, that is, QoS and mobility. Regarding the services supported by each protocol, the direction of SIP is quite different compared to the one of H.323. The former was designed as a general transaction protocol for setup and tear down of generic sessions, whereas the latter was designed with a specific view to multimedia applications, including telephony. Therefore, SIP provides for a broader spectrum of services than H.323. However, H.323 provides a more precise and detailed specification of voice and multimedia services. In regard to supplementary services, H.323 has the upper hand since SIP presents some drawbacks concerning the standardization of supplementary services. This also led to the conclusion that H.323 provides better interoperability and interworking with PSTN. Despite the fact that H.323 has been the industry's predominant standard for offering multimedia services over IP networks, with the advent of SIP-based networks it has become certain that none of the two protocols will dominate over the other. What is most likely is that the protocols will coexist over a longer time and this is the reason why interworking between these two protocols aroused researchers' interest [87,88].

As far as IAX2 is concerned, it is a new protocol and has not been compared to other signaling protocols as yet. IAX2's main advantage stems from its unified signaling and audio paths which permit it to transparently navigate NAT firewalls since only one port needs to be opened. Moreover, this nifty protocol is more bandwidth efficient than the other protocols, and bandwidth usage can be further reduced by IAX2's ability to aggregate RTP traffic. Notwithstanding these advantages, IAX2 still lags behind SIP and H.323 in the following areas. IAX2 is suited for carrying voice and video data, while SIP is able to transport nearly everything. H.323 can also be used for a multitude of services, although not being so broad in scope as SIP. What is more, in contrast to SIP and H.323, IAX2 is implemented only in some IP-phones, therefore, many vendors decide to use SIP or H.323 because all the phones in the market are compatible with these protocols.

So, which protocol is the right one to use? Unfortunately this question still concerns most of the people in this field and there is no simple answer. The decision on a protocol must be taken after a careful examination of the network where the protocol will be applied, its potential growth and several other factors [77]. For instance, H.323 is suited to midsize and large networks since it offers a great deal of scalability, reliability, interoperability and integration with PSTN. SIP can virtually be implemented in a network of any size due to its simplicity, scalability and low overhead. Regarding MGCP and H.248, as previously stated, these protocols work on a simplified, centralized model. They can be used for a Greenfield IP telephony deployment [77], however, we should always keep in mind that these protocols do not specify mechanisms for synchronizing call agents. Hence, they can form a part of large scale H.323 deployments, where terminals use H.323 for communicating with each other and with

the network, while these protocols can be used to control large gateways that interconnect the VoIP network to PSTN [89].

9. QoS via call admission control

The advent of IP networks has given service providers tremendous opportunity to offer both traditional services and a range of new services. However, providers need to find incentives for users to use and pay for these services. In this context, QoS represents an issue of utmost importance. VoIP poses substantial challenges, the most important one being perhaps resource management. As befits a subject of such importance, Call Admission Control (CAC) techniques for VoIP have been the subject of considerable study. The task of CAC mechanisms is to decide on whether a new voice flow can be admitted into the network with guaranteed QoS, without breaching QoS guarantees made upon the establishment of active flows. Before setting out the salient features of the CAC mechanisms that have been proposed in the literature for VoIP thus far, we shall first describe the issues that a CAC technique grapples with.

When no admission control is implemented, under congestion in the IP network all calls experience packet loss, thus voice quality deteriorates. It has been shown that the degradation of the quality perceived by users depends on several factors such as the voice codec, the packet size and the burstiness of the packet loss. It is indubitable that voice becomes unintelligible when packet loss is around 10–20%. Moreover, it has been reckoned that voice can tolerate as little as 1% packet loss. In addition, under congestion the end-to-end delay increases rapidly and so does delay jitter. Markopoulou et al. assessed the performance of VoIP in the US backbone networks [90]. Albeit these networks are usually over-provisioned, poor performance was observed in many cases. The ability of the current Internet to support voice communications was also evaluated in [91]. Statistics collected from international and commercial Internet paths divulged that the performance of VoIP is still some steps away from that offered by PSTN today. Thereby, a challenging task is to devise mechanisms to satisfy the packet- and call-level requirements of voice. Toward this end, several solutions have been proposed.

At the packet level either the Integrated Services (IntServ) model [92] or the notable Differentiated Services (DiffServ) model [93] can be employed to provide service guarantees. In the former model, service differentiation is focused on individual packet flows, thereby requiring end-to-end signalling (i.e., Resource Reservation Protocol – RSVP) for reserving resources along the path. However, this approach raises scalability issues and renders this model inappropriate for networks with a large number of nodes. In the latter model, service differentiation focuses on aggregates of flows. In other words, flows are categorized into service classes. At every node, each packet is treated according to its class. In this study we would rather not dwell upon these two QoS architectures nor accentuate the performance disparities between them since they have been exhaustively studied in the literature as yet. In brief, DiffServ presents clear advantages over the IntServ approach such as its salient scalability. Therefore, it seems to be the best candidate for providing QoS guarantees in VoIP systems since voice can constitute a traffic class itself. In [94], the DiffServ model was used to isolate voice traffic from bursty TCP traffic. Nonetheless, traffic overload can still occur due to the statistical behavior of voice or link failures. To surmount this problem, a selective dropping mechanism was employed to gracefully degrade voice quality during congestion. According to this mechanism, low priority packets are first dropped in order for high priority packet to be transmitted with a lower loss rate. Simulation results provided corroboration to the enhanced performance of that technique compared to the Best Effort model.

Another way to improve the performance of the DiffServ model, as well as that of IntServ, is by providing QoS support at the network layer as well. In the current Internet, routing is focused on connectivity rather than on QoS provision. Current Internet routing protocols are optimized for a single arbitrary metric in order to be highly scalable. Moreover, they do not take into consideration the availability of resources and the QoS requirements of each service class in the path selection process [95]. Consequently, VoIP flows, and in general real-time flows, may be transferred over paths that cannot support their QoS constraints. The problem of QoS provision at the network layer is called QoS routing, also known as constraint routing. The goal of QoS routing is to compute paths that can meet the QoS requirements of each service class, while maximizing the number of flows that can be admitted into the network [95]. To this end, QoS routing protocols aim at finding paths that satisfy multiple constraints. In the multi-constraint case, each link has multiple weights which can be classified as additive, multiplicative or concave [96,97]. Additive weights represent metrics whose impact on the end-to-end weight is additive. Such a metric is delay. A multiplicative path weight is the product of the link values along the path. An example of multiplicative weights is path reliability. Concave weights correspond to metrics such as bandwidth. Namely, the weight of a path is equal to the minimum link weight of the links that make it up.

Concerning VoIP, the metrics that affect the perceived quality of voice are delay, delay jitter and packet loss. Thereby, the routing protocol should compute paths that satisfy all the constraints related to these three metrics. Nonetheless, the cost to pay is an increase in complexity. The computation complexity is primarily determined by the metrics that are taken into consideration in the path discovery procedure. It has been proved that the complexity of finding a path subject to two or more independent additive and/or multiplicative constraints in any possible combination can be burdensome [98]. Thus, the only tractable combinations are those that consist of a concave metric and an additive or multiplicative metric. Delay and delay jitter are additive metrics, whereas packet loss can be easily transformed to multiplicative metric [98]. Therefore, complexity cannot be easily reduced. One solution that has been proposed to alleviate this problem is the use of a *single mixed metric*, i.e. a metric that is a combination of several metrics. However, a single mixed metric alone is not enough to determine if QoS requirements are

satisfied because some information is lost [98,97]. The Restricted Shortest Path approach represents another remedy that uses two additive metrics. In this case, the routing protocol first computes all paths that satisfy the constraint associated with one of the metrics and then selects among them the best one according to the second metric [99]. Another way to reduce complexity is to consider residual bandwidth and propagation delay as the constraints that must be satisfied. The first metric is concave, whereas the second one is additive. Hence, complexity is tractable in this case. Although this solution is not optimal, it constitutes a trade-off between complexity and performance [98], while the remaining requirements can be considered during the CAC procedure. Such a routing protocol was presented in [100]. The proposed protocol employs two metrics, bandwidth and the hop-normalized metric which is a normalized function of delay. The simulation results presented in that study proved that this protocol can satisfy the delay and jitter requirements of VoIP.

Scheduling mechanisms constitute another remedy to the problem of providing QoS guarantees. The basic function of the scheduler is to arbitrate between packets that are ready for transmission on the link. The First In First Out (FIFO) scheduling discipline is not only the simplest scheduling policy, but also the most widely deployed one in the Internet today. As its name suggests, a FIFO scheduler serves packets according to their arrival order. This scheduling policy does not provide any guarantees to end-users. Therefore it is not suitable for voice traffic. The Earliest Deadline First (EDF) is a dynamic priority scheduler with an infinite number of priorities [101]. Whenever a packet is transmitted, the scheduler searches the queue for the packet with the closest deadline. Weighted Round Robin (WRR) [102] is another scheduling policy that aims to provide QoS guarantees. It aims to give a weighted access to the available bandwidth to each class, ensuring a minimum allocation. The scheduler serves each class in a round robin manner according to the weights. If one or more classes are not using their full allocation, then the unused capacity is distributed to the other classes according to their weights. Another alternative to WRR is Weighted Fair Queuing (WFQ) [103]. WFQ and its variants aim to distribute available bandwidth over a number of weighted classes. WFQ effectively controls the ratio of bandwidth distribution among classes under congestion.

Although all the aforementioned mechanisms can improve the performance of VoIP systems, they do not always succeed in providing high-quality voice communications. Therefore, in addition to congestion detection/recovery approaches at packet level, it makes sense to use congestion avoidance approaches. One method that can be used to enable voice quality in IP networks is the so-called Capacity over-provisioning approach. This method is predicated upon the view that bandwidth is available in abundance; therefore the network capacity is always much beyond the traffic loads expected. Albeit this approach appears to be sufficient for the initial operation of the network, it is neither scalable nor viable in the long run [104]. Another approach is to allocate adequate slack capacity for voice calls, which in any case constitute a small fraction of the total network traffic. However, it is quite difficult to determine in advance the over-provisioning factors needed to meet the desired QoS [9]. Further, since no systematic procedures exist for this purpose, as the size of the network increases, so does the uncertainty about the accuracy of these factors. Therefore, admission control is also needed to provide QoS. Fig. 10 presents the architecture of admission control. The admission control unit makes the admission decision on the new call based on admission criteria, network state and flow information. Next, we succinctly summarize several CAC methods that have been proposed in the literature thus far. For the sake of presentation, the CAC techniques are classified into categories.

9.1. Per call parameter-based admission control

The most straightforward approach is to reserve bandwidth on a per call basis. Namely, at the initiation of each call a path should be established between the two far-off terminals and adequate capacity should be reserved along each link, that is, bandwidth is reserved on a hop-by-hop basis. This approach is based, to some extend, on the IETF's Integrated Services (IntSev) architecture and a protocol such as RSVP (resource reservation protocol) [105] can be used to support the associated per-call signaling. In brief, the originating terminal sends an RSVP message towards the terminating phone. This message carries information about the traffic characteristics of the session as well as information about the required QoS. At each node, RSVP attempts to make a resource reservation for the stream. If resources are not available, the session request is renegotiated or blocked. It should be noted that RSVP does not perform its own routing; instead it uses underlying routing protocols to determine where it should carry reservation requests. As the routing protocol changes paths to adapt to topology changes, RSVP adapts its reservation to the new paths wherever reservations are in place. Even though this method can in principle provide absolute QoS guarantees and attain efficient bandwidth utilization, it does not scale well in large IP networks because the state-of-the-art in IP technology is far from being amenable to support per-call bandwidth reservation. Current routers are not designed to handle large volumes of signaling messages. Thus, the signaling and processing at the intermediate routers impose a burden on their processing capabilities.

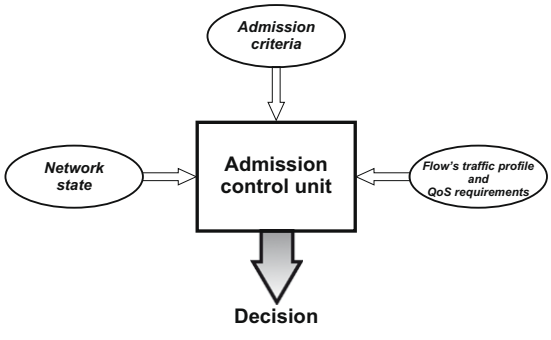

Fig. 10. Architecture of admission control.

Wang et al. proposed a framework to integrate utilization-based admission control with the existing VoIP systems [106,107]. The proposed QoS-provisioning system supports two types of QoS guarantees: deterministic and statistical guarantees. The CAC schemes that were examined were based on a predefined utilization bound. For each new call, as long as the sum of the used and the requested resources are not beyond the predefined threshold, which is computed offline and set at the configuration time, the service guarantees can be provided. Two kinds of utilization-based algorithms were tested. One that performs CAC based on the utilization of each link and another that performs CAC based on the bandwidth allocated to pairs of Gateways at the configuration time. When a new call is admitted into the network, the available bandwidth is decreased by the bandwidth requested by the call. The main feature of the aforementioned techniques is that the utilization threshold is determined and verified at the configuration time, so the real-time admission test is simplified to a utilization check which is performed at the admission decision making module.

A utilization-based CAC algorithm is also described in [108,109]. The authors of these studies took the DiffServ architecture as their base. As in [106,107], an ingenious calculation method is used offline in order to estimate for a given bandwidth allocation the delay bound for every class at each router. Upon the arrival of a new flow of the ith class, the admission control procedure only needs to check if the acceptance of the new flow will not result in exceeding the bandwidth allocated to this class on one of the links that make up the path.

A CAC scheme for provisioning QoS to voice traffic over DiffServ networks was also proposed in [10]. The proposed scheme relies on the extended Negligible Jitter conjecture. The CAC technique can be broken down into two modules. When a new call arrives, the first module ensures that the rate of the aggregate voice traffic only exceeds the capacity of the output link with a small probability. If the first check is successful, then the request for a new call is handled by the second module. The second module performs a numerical analysis to check if end-to-end delay and packet loss metrics of the voice flows are below the ones required for acceptable quality. The admission control is carried out at each router along the path based on parameters that describe the characteristics of each voice flow. If one of the checks along the path fails, then the call is rejected.

9.2. Per call reservation with path-based bandwidth allocation

Another effective strategy is to set up label-switched paths between VoIP gateways using MPLS (Multiprotocol label switching) and then reserve appropriate bandwidth along each path, which is called virtual trunk group (VTG) [110]. Then, the CAC scheme simply has to count the number of calls on each path or equivalently the bandwidth usage and block incoming calls when the path is fully utilized. Therefore, in contrast to CAC schemes of the previous category that are implemented locally at each router, in this case CAC can be implemented at the session manager which only has to keep track of the used and available bandwidth of each VTG. For the establishment of VTGs either the RSVP-TE (RSVP-Traffic Engineering) [111], an extension of the RSVP protocol, or the Constraint-based Routing Label Distribution Protocol (CR-LDP) [112] can be used.

One of the advantages of this approach is that it enables absolute QoS guarantees. In addition, the bandwidth allocated to each VTG can be dynamically changed (for instance, it can be reduced when the path is under-utilized). Notwithstanding these appealing advantages, this VoIP architecture is subject to some significant shortcomings. The first one is that maintaining an account of the ongoing calls is prone to a number of errors. It is quite difficult to predict the bandwidth requirements of calls with silence suppression. This may be compounded by codec changes amid ongoing conversations. Another drawback is that the number of paths to be set up grows as the square of the number of VoIP gateways. To circumvent this hindrance, the authors of [110] proposed a hierarchical architecture where gateways are divided into clusters and each cluster is managed by one control resource manager. These clusters (level-1 areas) form a new layer (level-2 area). Each cluster in the level-2 area is viewed as a single gateway and clusters are connected by VTGs in level-2 area. Thereby, any connection between two gateways that belong to different clusters goes over three VTGs: a VTG between the originating gateway and the core gateway (that is, the gateway used for connecting this cluster to other clusters), the VTG between the two core gateways of the corresponding clusters and the VTG between the second core gateway and the terminating gateway. Nonetheless, this hierarchical approach does not reduce complexity significantly.

9.3. Per call reservation with link-based bandwidth allocation

As mentioned earlier, the main pitfall of the per call reservation with path-based bandwidth allocation is the complexity associated with the management of paths' bandwidth. To overcome this shortcoming, Houck and Meempat proposed a CAC algorithm that is based on a link congestion metric rather than a path congestion metric [9]. That technique requires that label switched links are reserved. Nevertheless, instead of allocating bandwidth to each path, a certain aggregate bandwidth is reserved for voice in each link by virtue of a DiffServ class. Although not being identical, this approach shares many commonalities with the Bandwidth Broker architecture proposed in [113] for scalable support of guaranteed services. Therefore, it becomes possible to perform admission control by tracking the voice occupancy of each link rather than of each path. This means that the complexity is a function of the number of physical links instead of the number of paths. For the implementation of such an algorithm a new functional module is required to maintain a complete topology image and a consistent view of the usage status of the links in the network. This module is named call admission manager (CAM).

The algorithm proposed in [9] operates on a per call basis. Upon the arrival of a new call, the CAM opts for the best path for the particular source–destination pair. Then it

identifies the bottleneck link of this path and compares its occupancy to a threshold. Upon admission, the CAM updates the occupancy metrics of the links in that path. The occupancy metrics are also updated when a call is terminated. The main drawback of this technique is that it requires per-call communication between the CAM and the associated SoftSwitches, hence its scalability as the network grows can be questioned.

9.4. Measurement-based admission control

A traditional admission control approach consists in using measurements to determine the network state. There are several variations of this idea. Refs. [114,115] review some of the proposed measurement-based CAC algorithms. An advantage of this kind of method is that it makes no assumptions about the network state. In this subsection we delineate those measurement-based CAC schemes that are geared towards VoIP. The techniques are categorized according to the types of measurements they use.

9.4.1. Packet loss measurements

According to this approach a user wishing to establish a voice call first sends a preamble packet stream (such as ping) to the intended destination and measures the packet loss [116,117]. This method is usually referred to as probing phase. Probe packets are generated at constant rate and have lower priority than voice and data packets. There exist two variations of this approach: (i) the on-demand probing; and (ii) the constant probing.

In the on-demand probing method, the probing phase is invoked at the arrival of each call request [116]. Elek et al. also proposed a CAC scheme based on on-demand probing in [118]. The sender first sends a stream of low priority probe packets. The destination counts the received packets until the probing phase expires and then sends a report back to the sender. Based on the report the source decides on whether to accept or reject the new call. The main difference between the technique proposed in [116] and the one described in [118] consists in that in the former scheme probe packets are sent in both directions and each end node performs the admission test based on the measurements taken at its side. Therefore, a new call is accepted only if both tests are successful. A technique akin to the one presented in [118] was proposed in [119]. The new concept introduced in [119] is the probability-based admission decision. Specifically, a new flow is rejected with probability $p = 1 - f(x)$, where $f(x)$ is a monotonic increasing function of the predetermined admission threshold x. In that study the simple function $f(x) = x$ proved to perform fairly well. However, the selection of the optimal function constitutes an as yet open issue. An apparent weak point of on-demand probing is that it increases the call setup time. The latter can be reduced by decreasing the number of probe packets. However, this in turn renders the accuracy of the measurements questionable.

In the constant probing approach, probing is continuously conducted for each pair of end nodes (i.e., routers connected to a LAN or a VoIP gateway) [117]. Each one of the communicating end nodes maintains the latest packet loss rate. Therefore, upon the arrival of a new call, each one of the end nodes checks if the packet loss rate from the probe flow from the other end node is below a predefined threshold. The outcome of both tests should be successful in order for the call to be admitted into the network. Although this approach significantly decreases post-dial delay, it introduces extra traffic in the network, especially when the number of end nodes is large.

Another variation on obtaining measurements is to utilize the performance information signaled in RTCP packets of active calls [120]. Thus, there is no associated bandwidth overhead. The technique proposed in [120] operates in each VoIP gateway independently of the other gateways in the network. Moreover, statistics are maintained for each independent path between two end nodes. For each RTP stream, the destination gateway measures call quality statistics such as packet loss, delay and delay jitter. However, in the technique proposed in [120] only packet loss measurements are utilized. For each path the maximum observed packet loss ratio is measured based on 10 randomly picked flows (or all flows if less than 10), since it would be intractable to determine the maximum packet loss over all associated flows. The measured statistics are then sent back to the source gateway periodically in a special field within RTP packets or in RTCP packets. The reporting period is a parameter of the algorithm, referred to as CAC update interval. Based on the reported measurements, the CAC scheme simply compares the measured packet loss to a predefined threshold.

Jeske et al. proposed two CAC algorithms that utilize packet loss measurements in [121]. Based on packet loss rates measured by gateways, both schemes aim to predict the packet loss rate that future calls will experience. Two statistical models for predicting packet loss rates were examined. The first one was a simple exponentially weighted moving average (EWMA) model, while the second one was a sophisticated auto-regressive (AR) model. A comparison of these two models revealed that the former model performs fairly well compared to the latter one, while being much less complex. Both of the proposed CAC schemes make use of the EWMA model. The difference between them lies in the interface where packet loss rates are measured. One scheme uses measurements collected at the network layer, which represent packet losses due to network conditions, whilst the other uses measurements collected at the application layer at the time where the packet play-out is about to take place. Simulation studies proved the superiority of the latter scheme to the former one.

9.4.2. Delay jitter measurements

The study in [122] deals with admission control based on delay jitter measurements. Measurements are taken by the end nodes using probe packets. As in [116,118], the probing phase consists in the transmission of a number of low priority probe packets with fixed inter-departure time and the decision on whether to accept or reject a new call is made at the destination node. For each one of the received probe packets the destination node computes the time interval elapsed from the arrival of the previous

packet until the arrival of the current packet. This interval represents the variation in delay between consecutive packets and it should be within specific limits. If this check fails even for one of the probe packets then the call is blocked. The final outcome of the acceptance test is communicated to the source node by means of high priority feedback packets. If no feedback packet arrives at the source node before the expiration of the probing phase, the call is rejected. Compared to the schemes that measure the packet loss ratio of probe packets, this technique appears to be more efficacious on account of the fact that the performance of the former schemes depends on the buffer size of each router. Therefore, probe packets may be cast aside due to their low priority and not because of congestion.

9.4.3. Delay measurements

An alternative to call admission techniques based on either packet loss or delay jitter measurements is presented in [123]. The authors of that study proposed an algorithm that relies on packet delay measurements. Edge devices, such as media gateways, are shouldered the responsibility to detect congestion and perform admission control. Toward this end, each edge device measures one-way delay to the edge devices with which it communicates and keeps a record of the most up-to-date measurements. An independent record is maintained for measurements over control packets, which are assigned the highest priority possible and sent at a constant rate. Based on these records, the edge device can then detect congestion. In particular, based on measurements over control packets, the algorithm computes a delay threshold. If the measurements over voice packets are above this threshold, then congestion is implied and any new call request is rejected. Congestion is considered to be mitigated when all these measurements are below this threshold. One shortcoming of this technique is the difficulty in estimating the optimal threshold. In addition, in that study control packets were transmitted at a rate of 10 kb/s, thus making up for an overhead that cannot be overlooked.

9.4.4. Link utilization measurements

Per call reservation with path-based bandwidth allocation is not suited to large networks. Houck and Uzunalioglu showed that for many services, such as VoIP, a link-based bandwidth management approach based on measurements is more efficient, more scalable, easier to manage and more robust to unexpected changes in traffic demands [124].

In [9] an algorithm that is similar to the one described in Section 9.3 is proposed. However, the CAM now uses occupancy metrics that are based on measurements performed every T seconds. The update interval may be chosen so as to make the approach scalable for networks of any size. Nonetheless, in this case all the decisions made in the meanwhile are based on link utilization measurements which may be out-dated. The authors of that study claim that this drawback can be marginalized by maintaining slack capacities on the links. Obviously, there exists a trade-off between the duration of the update interval and the slack capacity.

Another CAC scheme based on link-measurements is proposed in [104,125]. The CAM uses SNMP (Simple Network Management Protocol) to receive occupancy metrics of voice traffic on each link periodically. Further, three voice priority classes are considered. When a congested link is detected, the CAM is in charge of forming a policy for all paths that use this link. The policy has the following form: "Accept all new calls between network ingress A and network egress B if link utilization is below a low threshold. Block x% of new calls if link utilization is between the low and a high threshold. Block all new calls if link utilization is over the high threshold". For a multi-class environment, a different pair of thresholds should be used for each class. The policy, however, does not apply to calls of the highest priority class. Calls of this class are always admitted into the network. What is of paramount importance for the performance of that scheme is the rules that govern the translation of a given link utilization into a blocking rule. To make the policy function more adaptive to traffic load, in both [104,125] a dynamic policy function was employed, that is, the percentage of new calls that will be accepted if link utilization is higher than the low threshold can be changed dynamically. Simulation experiments led to the conclusion that the best performance is attained when the CAM polls the IP routers every 30 s.

9.4.5. Delay and link utilization measurements

The measurement-based admission control approach advocated in [126,127] makes use of delay and link utilization measurements. To estimate delay, the queuing delay of each packet is measured. The average delay is then computed by averaging out the measured values over a window of T packet transmission units. As regards link utilization, it is extrapolated from measurements of the aggregate bitrate for each class over a period of S packet transmission units. Upon the arrival of a new call, the algorithm rejects the request if the sum of the flow's requested rate and current aggregate bandwidth utilization exceeds the target link utilization level or if the admission of the new flow would violate the delay bound. The algorithm operates on a per call basis, therefore a signaling protocol such as RSVP is required. Extensive simulation studies under a plethora of different network topologies and traffic models showed that this CAC scheme can attain high link utilization, while meeting delay requirements at the same time.

9.4.6. Link loading measurements (packet marking)

Gibbens and Kelly proposed another approach where packets encountering congestion are marked [128]. There exist several marking strategies. One consists in marking all packets that arrive in a buffer that contains more than k packets. In [128] an alternative marking strategy is described that aims to provide early warning of congestion. The gateway decides on whether or not to accept an arriving call according to its recent experience of marked packets. Although this mechanism does not measure directly the packet loss ratio, it is closely related to the schemes that do so in the sense that marked packets warn about the possible loss of packets.

9.5. Conclusions

Albeit parameter-based admission control can provide hard delay guarantees, it may over-commit the underlying network resources since it is based on the description of the traffic characteristics, hence aggravating network utilization. In addition, it should be noted that while some parameters can be easily specified (for instance, the peak rate), the actual average rate is difficult to estimate. To cope with the issue of low bandwidth efficiency, measurement-based admission control has been proposed. Measurement-based CAC techniques can only provide soft delay guarantees but they better reflect the dynamics of network status. There exist two flavors of this type of CAC strategies. The first one refers to those techniques that rely on measurements taken over probe packets, whereas the second one relates to CAC methods that capitalize upon measurements taken over packets of ongoing calls. The drawback of the former kind of schemes is the increase in call setup time that the probing phase induces, whilst the shortcoming of the second type of schemes is the dependency of the algorithm's performance on the update period of measurements since stale information may impair the algorithm's operation.

Notwithstanding these pitfalls, measurement-based techniques appear to be the most appropriate solution for VoIP networks since they can capture network dynamics and provide a good estimation of packet delay, delay jitter and packet loss. In [129], measurement-based admission control was compared to parameter-based admission control in the context of wireless IP networks. The effectiveness of the latter kind of admission control sharply deteriorates as traffic load increases and its performance is inferior to the one of measurement-based CAC schemes. However, one question still remains to be answered. Which kind of measurements is apt for voice traffic? Delay measurements, delay jitter measurements or packet loss measurements? Unfortunately, there is no clear-cut answer to this question. The nature of the system will probably dictate the type or the combination of measurements that will yield the best performance.

Moreover, as shown in [130], the performance of CAC schemes can be improved when they are combined with dynamic bandwidth allocation (DBA) schemes. DBA algorithms measure the amount of bandwidth that is used on a regular basis and reallocate it with the aim of increasing the number of calls that can be served by the network. In [130], the proposed DBA scheme simply decreases the capacity that is allocated to some calls (equivalently, voice is encoded at a lower bitrate) when congestion is detected. DBA schemes can also be very effective in wireless systems when acute variations in the signal power impose the use of a strong FEC technique, thereby decreasing the information bitrate. In that case, more capacity can be allocated to real-time services to make up for the decrease in information bitrate due to the ensuing redundancy.

10. Security

While most of the technical hurdles appear to have been overcome, security still constitutes a major concern. Security hinges on allowing only authorized users to make calls as well as ensuring that the content of a conversation is not eavesdropped. VoIP brought a new set of protocols that expose the network to the whim of technically savvy users. Several groups have pointed out several design and implementation flaws in VoIP's building blocks. In the following section we provide a cursory review of the main VoIP threats, and next, we dwell on the impact of security implementations on voice quality. For more details on security issues in VoIP systems, the voracious reader is referred to [131], which represents a voluminous treatise on this field.

10.1. Voice threat categories

Stanton in [132] classifies the threats to VoIP systems into six categories:

- *Denial-of-service (DoS):* This kind of attack involves flooding the network with spurious data, hence reducing its ability to carry voice calls. The majority of DoS attacks are based on exhausting VoIP servers' resources such as memory, CPU and bandwidth. An attacker can also flood a target VoIP system with a host of false service requests that will impede the system's ability to handle legitimate calls. A recent study revealed that SIP provides a broad spectrum of features that can be used to mount DoS attacks as well as countermeasures [133]. Interference can also stem form legitimate actions such as the downloading of large files. The problem can be alleviated by separating voice traffic from data traffic, either by giving priority to voice over data packets or by dividing the available bandwidth into two or more logical networks. Nonetheless, the incidence of this kind of attacks has so far been low.
- *Theft of service:* The aim of this type of attack is to make phone calls at someone else's expense without their permission. In VoIP systems, this can be the result of inadequate network security or infection of IP phones and softphones by malware. Security against these attacks can be attained by applying strong authentication solutions and installing antivirus solutions to protect IP phones.
- *Telephone fraud:* This category refers to the case where telephone fraudsters direct calls to premium rate numbers without the user being aware. This can be achieved either by malicious software (often called dialers) or by devices that can be attached to an organization's network. In private networks this threat can be dealt with by limiting the range of numbers that a user can call. As far as measures against malicious software are concerned, they can be curbed using special software to detect any applications that appear to be rogue dialer software.
- *Nuisance calls:* Spam e-mails are a real nuisance to anyone nowadays. While rare at present, Spam over Internet Telephony (SPIT) is a particular worry for the future. SPIT is akin to e-mail spam in that senders generate an abundance of calls for marketing purposes or to disrupt user's activities. However,

voice is much more difficult to protect than e-mail since the decision on whether a call is spam or not should be made within call setup time. A solution was proposed in [134] according to which call patterns are continuously monitored and spam calls are determined based on these patterns.

- *Eavesdropping:* The aim of eavesdropping is to listen in on calls. There exist several tools that can sniff IP packets and then stitch them together into an audio file, using a tool such as VOMIT (voice over misconfigured internet telephone). This problem occurs only when voice and data packets are transmitted over the same logical network, for example in the public Internet. End-to-end encryption techniques[18] can be used to protect the confidentiality and integrity of the information carried by IP packets. However, encryption also makes harder for law enforcement agencies to gather evidence against criminal cases. While the government's right of wiretapping is being heavily debated, it is also universally agreed that safety gets higher precedence over privacy. This paramount issue is discussed in [135], where the VoIP security issues are reviewed with the law enforcement agency perspective.
- *Misinterpretation:* This type of threat is related to attacks that aim to elicit confidential information from users. In this case, attackers can gain the ability to masquerade as a legitimate organization. Users that hide behind acceptable addresses can be clogged up by anti-spoofing packet filters.

10.2. Security mechanisms

Recommendations on how to address security issues in VoIP systems are spelled out in [136,137]. The first step to secure a VoIP network is that of identification and authentication. The fact that there does not exist a physical association between end devices and phone numbers makes identification rather difficult, though not impossible. A potential mechanism for providing identity management and authentication is the use of digital certificates. In [136] four different frameworks are compared and the merits and demerits of each one are pointed out. Once authentication has been carried out, authorization becomes the next step in the security process. Authorization is the process during which the actions that a user can undertake are determined. Several techniques can be used for this purpose and three of them are discussed in that study. Perhaps, the most important requirement for a secure environment is confidentiality and integrity. The former refers to ensuring the privacy of information exchanged, whereas the latter relates to ensuring that this information is not tampered with while being in transit. A number of approaches have been proposed thus far that can be employed to provide confidentiality and integrity in VoIP networks, the most prominent ones being IPsec (IP security) and secure RTP (SRTP). IPSec is a framework for securing IP communications by encrypting and/or authenticating each packet. It can operate in two modes: transport mode and tunnel mode. In the former mode, only the payload of the packet is encrypted, therefore, it is suited to host-to-host communications over a local area network (LAN). In the second mode, the entire IP packet is encrypted and must be encapsulated in a new packet so that it can be routed. This mode is used for network-to-network, host-to-network, or host-to-host communications over the Internet. SRTP defines a profile of RTP, intended to provide encryption, message authentication and integrity, and offers protection to the RTP data in both unicast and multicast applications [138], and is supported by both H.323 and SIP. Moreover, there are ongoing efforts in both ITU and IETF that aim to enhance H.323's and SIP's ability to secure connections.

Another well-known security mechanism that has served data networks well is the use of firewalls and NATs. However, as mentioned earlier, the obfuscation through which NAT works poses new challenges in establishing VoIP communications. Another significant issue addressed in [137] is that of availability. DoS attacks can cause lost of availability. Authentication and identification can be used to thwart this type of attacks. Notwithstanding, in the case of distributed DoS (DDoS), there is no clear-cut solution.

Solutions to three of the most important threats are discussed in [139]. In particular, the authors of that study elaborate upon three of the most daunting challenges in the context of VoIP security, that is, secure traversal of firewalls and NATs, detection and alleviation of DoS attacks and security against eavesdropping. The interested reader is referred to [139] for more details on these solutions. Although the potential voice threats have been identified and analyzed in many studies, there exists only a handful of tools that can be used to perform advanced security assessment tasks. The authors of [140] developed such a tool that is capable of evaluating basic security levels for a VoIP network.

10.3. Security over wireless links

VoIP over wireless links constitutes a daunting challenge in terms of security vulnerabilities since aside from all the issues that arise in the case of wired VoIP, the wireless link should be secured as well. In wireless systems the signal is transmitted over the air, so it is physically much easier to be intercepted by attackers. Moreover, attackers can gain illegitimate access to the network or can masquerade as legitimate access points, deluding users into connecting to a rogue access point. Therefore, in addition to the aforementioned security measures, wireless VoIP also requires some additional measures.

The Wired Equivalent Privacy (WEP) algorithm represents one of the first attempts to secure IEEE 802.11 wireless LANs (WLANs) [141], introduced in 1997. However, WEP hinges on a static encryption key and it has been proven that the WEP key can be determined if a number of packets are sniffed [142]. Nowadays there exist several software tools that can crack the key in as little time as 1 min or even less. Owing to its vulnerabilities WEP was

[18] With end-to-end encryption, the communicating entities initially exchange a secret key pair that they will be using to encrypt data. The larger the length of the key, the harder for an intruder to decipher the data.

superseded by the Wi-Fi Alliance's[19] solution, the Wi-Fi Protected Access (WPA), in 2003. In comparison with WEP, WPA uses a stronger encryption algorithm called Temporal Key Integrity Protocol (TKIP), which is still based on the same encryption algorithm as WEP, but is uses a longer key which is different for each packet. Moreover, it makes use of the IEEE 802.1X standard for user authentication [143], which ties to framework of the Extensible Authentication Protocol (EAP) [144]. In other words, the authentication process makes use of EAP messages.

One year later after the introduction of WPA the IEEE 802.11i was ratified with the aim of providing improved security in WLANs [145]. WPA can be viewed as a subset of IEEE 802.11i since it relies on several mechanisms that are used in the latter standard. The reason that compelled Wi-Fi vendors to embark on its development was that WEP was insecure and the industry could not wait for the 802.11i standard to be complete. Therefore, the two standards are similar. However, in addition to WPA's security mechanisms, 802.11i, which is also known as WPA2, includes support for the Advanced Encryption Standard (AES) [80]. In particular, it employs the Counter Mode with Cipher Block Chaining Message Authentication Code Protocol (CCMP) which uses the AES algorithm and is considered stronger than the RC4 algorithm that TKIP uses.

10.4. Security's impact on QoS

Several solutions to enhance security in VoIP systems have been proposed thus far. However, delivering voice packets to their destination undamaged is not sufficient. Voice has some stringent constraints in regard to end-to-end delay and delay jitter, therefore voice quality should be considered in tandem with security, for extravagant, high-security mechanisms can tantamount to breaches when certain aspects of security are disregarded. Encryption algorithms apply cryptographic functions to the packets' payload and build new headers. Hence, they introduce two kinds of delay which correspond to the amount of time needed for the encryption/decryption of voice packet. Considering that in most cases the payload of voice packets is small, this delay overhead for each packet can be detrimental to voice quality. The stronger the encryption that an algorithm provides, the greater the delay that it introduces [146]. In addition, in certain cases, encryption hides the QoS information required by the intermediate nodes to provide better service to the voice traffic [135].

Concerning the delay associated with encryption, Barbieri et al. pointed out two mitigation techniques [146]. The first one is the reordering of packets based on their QoS requirements before they are inserted in the cryptographic scheduler. This scheduler usually operates in a FIFO manner. Therefore, in many occasions voice packets are delayed because this scheduler processes larger packets. The second solution is to use hardware-implemented AES encryption, which is much faster than the software implementations of complex encryption algorithms.

[19] The Wi-Fi Alliance is a global, non-profit industry association of more than 300 member companies devoted to promoting the growth of wireless Local Area Networks (WLANs).

The authors of [147] studied the impact of encryption algorithms on the quality of voice calls in WLANs (based on IEEE 802.11b) and WPANs (based on Bluetooth). To this end, the E-Model was employed and the results were converted to MOS rating. The encryption algorithms that were tested were the Advanced Encryption Standard (AES) [80] and the Triple Data Encryption Standard (TDES) [148], both with 192 key length. Without encryption, MOS in both scenarios was around 4.25, while when encryption was employed, it was lower than 4. The TDES algorithm proved to have more detrimental effects on voice quality than the AES algorithm.

In addition to the noxious effect on end-to-end delay, encryption also has a deep impact on effective bandwidth. Encryption algorithms add headers to voice packets, thereby increasing the ratio of header size to payload size and reducing effective bandwidth. The effect of IPsec on effective bandwidth was addressed in [146]. IPsec provides for encryption and authentication at the network layer. Depending on the encryption technique employed, the length of the packet headers can be increased by up to 50 octets. Experimental results showed that the headers added to each packet due to IPsec decrease the effective bandwidth up to 63%.

Except for the increase in end-to-end delay, security mechanisms also increase call setup delay due to the identification and authentication mechanisms they employ. The effects of security measures on call setup delay, end-to-end delay, packet loss and delay jitter were investigated in [149]. A SIP-based architecture was considered and three different scenarios were examined. The first one was a network without any security measures, the second one was a network with standard firewall filtering, while the third scenario considered a virtual private network (VPN) that interconnected the communicating terminals. On the one hand, experimental results showed that packet loss and delay jitter are barely affected by security measures. On the other hand, they also disclosed that end-to-end delay and call setup delay increase in the presence of security implementations. Specifically, VPNs have the greatest impact on these two meaningful performance indicators. Concerning encryption, two techniques were assessed, i.e., AES and TDES. The former is considered to provide higher security, while the experiments conducted in that study proved that it performs similar to the TDES in terms of packet delay and call setup delay.

Fiedler et al. addressed the impact of firewall algorithms on voice quality [150]. They showed that as the number of the rules that the firewall depends on increases, throughput decreases, which translates in increased delay. The reason behind this trend is that searching for matching rules becomes more complex as the firewall ruleset grows. Possible solutions to this problem were discussed in [151]. One solution consists in designing more intelligent protection algorithms that would generate as few firewall rules as possible without compromising on the security level. However, the downside of this approach is that the complexity of the detection algorithm would increase. The authors of that study focused on another potential solution. Specifically, they proposed an algorithm that assigns a value to each rule.

This value shows the significance of the rule and is associated with the number of times the rule has been matched. The algorithm drops insignificant rules. Moreover, single-mapped rules, such as a IP addresses, are merged into a multi-mapped rule in order to decrease searching time.

The study in [152] investigated the impact of SRTP on voice quality. The performance indicator that was used to gage voice quality was MOS. Experimental results proved that when SRTP is enabled, MOS slightly deteriorates since delay increases. This translates in that if a target MOS must be met, either network delay or packet error rate must be reduced in order to compensate for the delay that SRTP introduces.

The impact of security measures on voice quality in 3G networks was investigated in [153]. In more detail, the effect of IPSec on delay, delay jitter and packet loss was evaluated for two voice codecs. One with 32 kb/s rate and another with 64 kb/s. Simulation results led to the conclusion that the integration of a security algorithm with VoIP has a great impact on the performance of VoIP in a 3G network. IPSec increases delay and delay jitter due to the significant increase in packet size caused by the additional header for security.

The impact of security mechanisms on another type of wireless systems was assessed in [154] where WPA and IPSec with the 3DES encryption algorithm were compared in the context of an IEEE 802.11b wireless network. Moreover results were provided for two different voice codecs; G.711 and G.729. Simulation results revealed that IPSec has a less severe impact on voice quality than WPA in terms of MOS, delay and packet loss regardless of the voice codec employed. The impact of WPA on voice quality was also evaluated in [155]. Moreover, in that study WPA was compared to WEP in terms of throughput in an IEEE 802.11g WLAN. It was shown that both mechanisms do not decrease throughput. Further, the effect of WPA on voice traffic was examined. Experimental results divulged that WPA does have an impact on voice quality and specifically on delay jitter only during call handover, while delay, throughput and packet loss remain unaffected.

10.5. Conclusions

Security is not an end game, but instead it is an evolving process. Network management requirements keep changing in order to accommodate new features. Even though the new component can be geared toward the specific needs of security management, they may well lead to security loopholes. Therefore, businesses and institutions should continuously reconsider their security framework, identify the nature and magnitude of new vulnerabilities and determine new courses of action. Even with best planning, it is inevitable that organizations will be attacked at some time, and unfortunately, the correct security strategy is neither a "one size fits all" solution nor a static implementation. The best way to face this harsh reality is to be in a position to isolate attacks as quick as possible.

11. VoIP over satellite links

So far we have surveyed many different areas of VoIP systems. Although the presentation was made in a general context, usually when one refers to VoIP, terrestrial systems spring to mind. This section aims to shed light to the potential of providing VoIP over satellite links. Fourth generation (4G) systems can be viewed as an ultimate amalgamation of existing and wireless fixed and mobile systems. Satellite systems can be instrumental in this future network infrastructure beyond any doubt, playing a multifaceted role. Consequently, integration of voice and data services is inevitable in satellite links as well. In fact, several satellite internet service providers have already expressed their interest to include VoIP as one of their service offerings.

While the service-related aspects of VoIP are the same in both terrestrial and satellite networks, transport-related issues are quite different. The most important factor that limits the performance of satellite systems is the propagation delay of satellite links that makes the provision of real-time service over this kind of links easier said than done. In the following section we address all the issues that are closely related to the performance of VoIP over different kinds of satellite links along with potential remedies.

11.1. Voice codec selection

The prime limitation of satellite systems is the longer propagation delay compared to terrestrial networks, which translates into degeneration of voice quality. A brief comparison of three well-known voice codecs was carried out in [156] in the context of VoIP over DVB-RCS (i.e., the standard proposed by ETSI for the return channel in satellite networks). The three codecs that were compared against each other were G.711, G.729 and G.723.1. Simulation results divulged that while the first two codecs exhibit acceptable performance, the performance of G.723.1 is poor on account of its high algorithmic delay. Using the E-Model, Janssen et al investigated the performance of four different voice codecs over satellite links and estimated the maximum delay that can be tolerated in each case [157]. Specifically, the G.711, GSM-EFR, G.729 and G.723.1 voice codecs were tested under various conditions, such as different echo levels, different packet loss rates and various packet sizes. It was shown that G.711 outperforms the other codecs since it can provide acceptable quality, while tolerating longer delay at the same time. Nevertheless, the penalty of G.711 is its high bandwidth consumption. When it comes to GEO satellite systems, a good candidate is also GSM-EFR because it allows for delays as long as 270 ms. However, as packet loss increases, the maximum delay that can be tolerated by each codec decreases. Based on these findings, the authors of that study concluded that Low Earth Orbit (LEO) constellations represent the best candidate for providing high-quality internet telephony over satellite systems.

11.2. Header compression over satellite links

Header compression techniques can be viewed as a stepping stone to efficient bandwidth utilization, which is of paramount importance in satellite networks. Despite the fact that header compression schemes have been the subject of significant research efforts in the context of terrestrial wireless systems, their performance over satellite links has never been investigated. However, the characteristics of the satellite channel may impair the performance of these schemes. Although some of the header compression schemes provide mechanisms to cope with the problems that arise in this occasion, all of them have been developed having terrestrial links in mind. The literature lacks a comparison among the three well-known techniques and especially between ECRTP and ROHC since these two have been devised for wireless links. In general, ROHC is expected to perform better than ECRTP. Nonetheless, it is questionable if the increase in the performance that ROHC provides can make up for its increased complexity compared to that of ECRTP. Moreover, the performance disparities among the three modes of ROHC needs to be investigated since the unidirectional mode is much less complex, while its performance may be close to the one of the two bidirectional modes owing to the long propagation delay of satellite links.

11.3. Signaling protocols in satellite systems

The hostile environment of satellite systems has an impact on the selection of the signaling protocol as well. The long propagation delay of satellite links along with the centralized nature of this kind of systems take a heavy toll on the time interval needed for the establishment of a call. In GEO satellite systems, a message needs roughly 270 ms to reach its destination if the satellite has on-board processing capabilities (i.e., switching can be performed onboard the satellite), while in the case of bend-pipe systems (i.e., the satellite acts as a simple relay and its only role is to amplify the signal before transmitting it onto the downlink) terminals communicate through a gateway station which translates in a delay around 500 ms. The latter case is accustomed in most of the current satellite systems, while all the future satellite networks are expected to fall into the first case. Furthermore, bandwidth in satellite systems is a scarce resource and signaling overhead should be as low as possible. All the signaling protocols that were discussed in Section 8 can be employed in satellite networks. However, their performance will not be the optimum since they have been developed having terrestrial systems in mind, where the problem of the long propagation delay is not acute.

Fig. 11 illustrates a general architecture of a GEO satellite system that provides VoIP service to its users. For the sake of simplicity, a satellite with on-board processing capabilities has been considered since its enables direct communication between two users. In that case, signaling messages are propagated to the gateway station which contains a gatekeeper that is in charge of processing signaling information. The gateway station also interconnects the satellite network with the Internet and PSTN. Concerning the interconnection with PSTN, the architecture is similar to the one depicted in Fig. 5. Namely, a VoIP Gateway is utilized.

The issue of signaling protocols for IP telephony over satellite links has been, nevertheless, disregarded until recently since IP telephony was not included in the bucket of services that most satellites systems provided. In the literature, there exist only a handful of studies that touched on it, with most of them being focused either on H.323 or SIP. Ott et al. investigated the performance of H.323 and SIP over long satellite links and came to the conclusion that both these protocols can be used in satellite systems [158]. Moreover, in the experiment they conducted H.323 proved faster than SIP in terms of call setup delay. SIP and H.323 were also qualified as suitable for satellite systems in [159]. The H.323 protocol was also adopted for the experiments conducted in [160–162], the focus of these studies were not on the performance of this protocol though. On the other hand, SIP was adopted in [163,164]. It is possible, however, that amendments may be required at other layers in order to reduce its call setup delay. With this end in view, Kueh et al. proposed some modifications to the radio link-layer of S-UMTS with the aim of decreasing the call setup delay of SIP [164]. The authors of [165] state that MeGaCO/H.248 may also be a good candidate

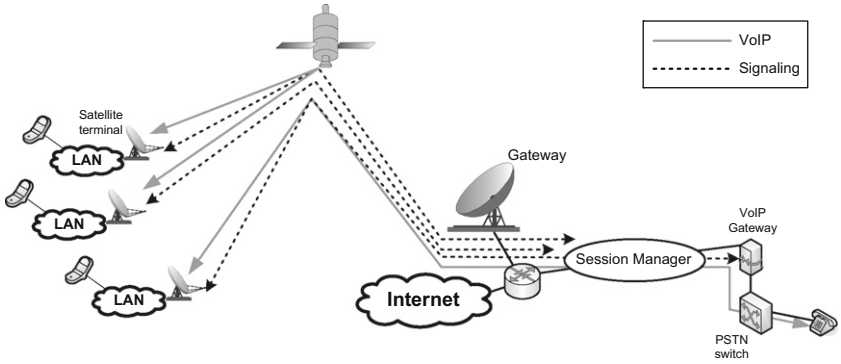

Fig. 11. Generic VoIP architectural solutions.

for satellite systems given its compact text messages (recall that in MeGaCo/H.248 several transactions can be grouped in one message). Nonetheless, this protocol does not constitute a complete system since it manages only gateways, and therefore, another protocol is required between gateway controllers. The study in [158] also concluded that the performance of MeGaCo/H.248 over satellite links would be poor on account of the fact that this protocol is highly interactive since it has been primarily conceived for the creation of terrestrial backbone voice networks. Concerning the IAX2 protocol, it may hold some appeal since it is less bandwidth-consuming in comparison with the other signaling protocols. Nonetheless, IAX2 is still far from being accepted by VoIP vendors. Last but not least, it is also highly likely that satellite vendors use proprietary protocols geared towards the limitations encountered in these networks.

11.4. VoIP over GEO satellite systems

VoIP over GEO satellite systems constitutes a daunting challenge. There are two reasons that make internet telephony service providers hesitant to venture on providing voice services via this kind of systems. The first one is the long propagation delay that characterizes this type of satellite links, while the other one is related with the packet loss due to fading. A feasibility study on the use of GEO satellite systems for the provision of internet telephony was presented in [160]. The performance of the network was evaluated under various bit error rates and link loading conditions. Moreover, two voice codecs were compared, namely G.723.1 and G.729. The performance indicators that were used for the assessment were IP packet loss, voice frame loss, bandwidth efficiency and packet delay. For the needs of the experiments a testbed was used and the results gathered proved that GEO satellite links represent a robust medium for transporting VoIP traffic. In addition to Nguyen et al. [160], a small portion of the results stemming from the European project VIP-TEN (Validation of IP-Telephony over EuroSkyWay Network) were presented in [166]. Specifically, delay and delay jitter measurements were presented, leading to the conclusion that VoIP over GEO satellite links is a viable alternative to terrestrial networks. Similar conclusions were also drawn in [167] where two LANs were connected via a satellite link provided by the GEO satellite SESAT. The average delay jitter of voice packets was around 5 ms and packet loss was fractional, while voice quality was high regardless of the traffic load carried over the link. The ability of the INMARSAT's Mobile Packet Data Service (MPDS) to support VoIP was investigated in [168]. The authors of that study collated results obtained via field tests, simulation experiments and a laboratory experimental platform. Ample experimental evidence led to the conclusion that MPDS can provide cost-effective voice services in oceanic, offshore and remote regions at the expense of increased, yet acceptable, latency.

Another issue that arises in the case of GEO satellite systems is bandwidth management. Bandwidth constitutes a commodity at a premium, and thus, bandwidth utilization needs to be optimized. To this end, the use of CAC and dynamic bandwidth allocation (DBA) schemes is deemed necessary. Both CAC and DBA decisions are taken at a designated gateway station, the Network Control Station (NCC). In the case of satellite with on-board processing capabilities, these decisions can also be taken on-board the satellite.

As far as CAC is concerned, both IntServ and DiffServ models have been studied for GEO satellite systems. Although the DiffServ model has an edge on the IntServ model in terms of scalability, it is reckoned that the latter model can serve satellite systems (and in general wireless systems) well since the number of connections that these systems handle is much smaller than the one in terrestrial wireline networks [169]. A review of CAC schemes for GEO satellite systems is provided in [169]. The workhorse of most algorithms consists in the concept of the *excess demand probability*, that is the probability that a given number of calls will request more bandwidth in a future time instant. The decision of the CAC algorithm takes account of this probability as well as traffic descriptor parameters. It should be noted, however, that all the algorithms have been devised having video service in mind but they can handle VoIP traffic as well.

Another issue of utmost importance is that of dynamic bandwidth allocation. Although the role of CAC is vital, it cannot always guarantee the fair allocation of bandwidth. This is the role of DBA algorithms. DBA schemes adjust the amount of bandwidth that is allocated to each call over time in order to optimize the overall QoS provided by the system. On the one hand, these schemes allow bandwidth to be saved, whereas, on the other hand, they usually increase access delay, i.e., the time interval elapsed from the arrival of a packet at a terminal's queue until its transmission. The impact of bandwidth on demand mechanisms on voice quality is addressed in [170], while an architecture of a GEO satellite system that is able to deliver VoIP services over the DVB-RCS standard is delineated in [171], where the NCC is responsible for handing out capacity to terminals. A CAC and DBA algorithm for VoIP traffic over satellite links is proposed in [172]. The rationale behind the proposed scheme is the exchange of information among the application, network and MAC layers. The rate of the voice codec is adjusted taking information about the network state into account. When there is no available resource to support a session at the lowest possible coding rate, then the algorithm attempts to decrease the rate from the active VoIP sessions on a round robin fashion, until there is sufficient bandwidth. The cross-layer approach is also adopted in [173] where it is shown that adaptive speech coding can also improve the performance of applications running in parallel with VoIP. This is achieved by dropping the speech coding rate when a certain level of congestion is detected. Moreover, the authors of that study concluded that the level of congestion can be retrieved directly from the MAC layer instead of the network layer because the latter cannot detect variations in available bandwidth due to the high propagation delay.

One important consideration specific to future GEO satellite systems is also queuing delay onboard the satellite. All the upcoming GEO satellite networks will make use of satellites with onboard processing capabilities in order to implement forward error correction and enable one hop

communications. In this case, a satellite can be modeled as a non-blocking switch that interconnects different input and output links. Each output port can, in turn, be modeled as a single server queue with finite buffer capacity. Buffering is necessary since different flows content for the same resources and scheduling policies that have been proposed for the terrestrial Internet can be applied to satellite systems as well in order to provide QoS to specific types of flows. Moreover, when the satellite interconnects two terminals that reside in terrestrial networks, then the NCC can decide on the best "landing point" so as to minimize end-to-end delay, delay jitter and packet loss. The decision can be based on information about the congestion of specific links in the terrestrial network [174].

11.5. VoIP over LEO satellite systems

Even though the interest in LEO satellite systems dwindled away on account of the limited commercial success of Iridium and Globalstar, the appealing features that this type of systems are endowed with, such as low latency, global coverage and the ability to communicate with handheld devices, along with the current trend toward the migration to all IP-based services open new opportunities to them, especially with regard to real-time services. Of the multitude of services that LEO satellite networks will be called to support, VoIP represents the most important one.

In the case of LEO satellite systems, end-to-end delay depends on the location of the communicating terminals. The propagation delay of up- and down-links is roughly 25 ms, while the propagation delay of inter-satellite links depends on the system's design and the position of the satellites in the orbit (usually it is in the range of 5–20 ms). Nonetheless, due to the movement of satellites with respect to the Earth's surface, a user must be handed over between contiguous satellites or cells. The handover of a call constitutes a daunting challenge in LEO satellite systems and many studies have grappled with it, mostly for the case of voice traffic. Two succinct reviews of handover strategies can be found in [175,169]. In this type of networks the issues of CAC and handover are interwoven. The usual approach is to admit a new call into the network only if the system can guarantee that this call will not be forced into termination during the first handover. Concerning subsequent handovers, all the algorithms that have been proposed in the literature aim at reserving bandwidth at the next cell before the handover occurrence. The time instant at which the reservation request is dispatched depends on the algorithm. Dropping probability, that is, the probability that an ongoing call is dropped, is dependent on the time interval between the time instants of the dispatch of the request and the handover occurrence. The longer that interval, the smaller this probability. Nonetheless, any effort to reduce dropping probability has an impact on blocking probability since bandwidth is reserved for ongoing calls against new calls. Concerning voice calls, this probability is regarded much more irksome for the users than the blocking of a new call. Therefore, most studies in that field seek for the optimal trade-off between blocking and dropping probabilities.

In this type of satellite constellations, the issue of dynamic routing also arises, and what is more, it plays a pivotal role in providing QoS guarantees. Routing data from the source all the way to the destination is a challenging task. Although routing techniques for non-GEO satellite systems have been the subject of extensive study, leveraging features of existing Internet routing algorithms and taking the predictable changes of the network topology into account, most of the proposed techniques do not provide for quality of service in the case of voice traffic. In [176], a routing scheme for a hybrid LEO–MEO (Medium Earth Orbit) satellite constellation is proposed and evaluated. Voice packets are classified at the source satellite (namely, the first satellite that packets traverse) into two categories depending on the propagation delay between that satellite and the destination satellite. For the estimation of the propagation delay the Dijkstra shortest path algorithm was used. If the estimated propagation delay was shorter than a predetermined threshold, the packets were delivered to the destination satellite through the LEO layer, otherwise they were routed through the MEO layer. The rationale behind this approach is to evenly distribute traffic load, thereby precluding excessive delay and delay jitter values. Furthermore, the performance of that hybrid system was shown to improve when satellites were able to implement weighted round robin scheduling.

The issue of routing in multi-layered satellite architectures was treated in [177,178] as well. In those studies a satellite constellation comprising an Iridium-like LEO system and three GEO satellites was considered. GEO satellites served the purpose of disseminating routing tables to LEO satellites, hence diminishing routing overhead. A multi-constraint routing scheme was proposed, according to which the process of computing the shortest path is dependent on available bandwidth, bit error rate and propagation delay of links. Additionally, the Multi-Protocol Label Switching (MPLS) QoS architecture was employed to shield voice traffic from bursty data traffic by assigning different paths to voice and data traffic. Simulation results documented the good characteristics of this routing algorithm in terms of end-to-end delay, delay jitter, packet loss and throughput.

11.6. Security in satellite systems

Apart from all the security issues that have been pointed out thus far, two other major issues arise when it comes to security in satellite systems. The first one is related to multicast communications while the second one is security at the link-layer. Even though the IPsec suit of protocols ensures the confidentiality and integrity of IP packets, should it be used for multicast communications, it would result in a considerable waste of bandwidth since. IPsec has been geared towards point-to-point communications and is not multicast aware, therefore, each link should be secured as if it was a unicast link. Moreover, the nature of satellite systems imposes some constraints on the security protocols. First and foremost, the protocol should add minimal overhead in order to ensure good bandwidth efficiency. Moreover, it has to be simply enough to reduce the requirements in processing power and robust to the loss of a few packets [179]. Oetting and King in [180]

addressed the first issue and proposed the use of header compression schemes and/or the inclusion of more than one voice frame in each packet. The European project SatIP6 developed a protocol termed SatIPSec (Satellite IPSec) [181] that meets the aforementioned constraints and ensures security at the network layer. SatIPSec builds upon IPSec extensions and aims to offer transparent and secure unicast and multicast IP transmissions in satellite networks. To this end, it hinges upon the work carried out by the IRTF SMuG (Internet Research Task Force – Secure Multicast Research Group), which was superseded by the GSEC (Group SECurity) group, and IETF MSEC (Multicast SECurity) groups. The main principle behind SatIPSec is that of centralized management, namely, it works on a server–client basis. A server configures each unicast client while in the case of multicast clients, all clients are simultaneously configured by the server.

As far as security at the link-layer is concerned, it is deemed essential in order to prevent an intruder gaining information through knowledge of the identity of the communicating parties and their traffic characteristics [179]. The link-layer security constitutes an additional security mechanism to the one at the transport layer and provides additional link confidentiality and receiver identity hiding over the transmission link. Current work in progress in the European funded project SATSIX aims to adapt SatIPSec to provide link-layer security [182,183].

12. Conclusions and future perspectives

Users and service providers are lured by the voices that praise countless benefits and new business opportunities, but can VoIP compete with traditional telephony systems? Albeit there is no clear-cut answer to this trite question, it is indisputable that Internet telephony is endowed with some outstanding characteristics and has the potential to revolutionize telephone communications. VoIP nowadays enjoys the fruits of labors during the past few years and it can be considered a mature technology anymore. Moreover, behind the concept of a new broadband network that will be the amalgamation of existing and emerging fixed and mobile networks lies the need for network operators to provide new broadband services, as well as the desire of customers to be able to have access to their services from anywhere. This concept is termed Next Generation Network (NGN). ITU-T defined an NGN as a packet-based network able to provide telecommunication services and able to make use of multiple broadband, QoS-enabled transport technologies, and in which service-related functions are independent from underlying transport-related technologies. A standardized NGN architecture is the IP Multimedia Subsystems (IMS), which was defined by the ETSI and 3GPP. IMS will work with any fixed or wireless network based on packet-switching, including GPRS, UMTS, WiMAX, DSL, etc. Furthermore, IMS builds on SIP in order to ease the integration with the Internet. Therefore, VoIP's future appears bright since a terminal only needs to support IPv6 and SIP. However, traditional telephony systems, H.323 and other VoIP systems can also be integrated with the IMS network through gateways. Notwithstanding the advantages of VoIP, one of the problems that still remains is the provision of QoS guarantees to voice communication over IP networks. This problem stimulated the development of novel mechanisms with the aim of improving speech quality.

In this article we surveyed all the aspects that have the greatest bearing on voice quality. Beginning with the QoS requirements of voice and methods that can be employed to evaluate the performance of VoIP systems, we continued with a survey on voice codecs and header compression techniques. Instead of glossing over these paramount topics, we provided an exhaustive analysis and comparison of different voice codecs and header compression schemes, touching also upon voice activity detection and packetization issues. We then turned our attention to signaling protocols and continued with a comprehensive study on CAC schemes tailored for the needs of VoIP. We also shed light to the meaningful issue of security in VoIP networks. It is concluded that significant strides have been made since the inception of VoIP and Internet telephony will soon be able to provide speech quality at a similar level to the one of PSTN. Last but not least, we discussed the potential of satellite systems to provide VoIP. Satellite systems present some appealing features, such as global coverage and their intrinsic capability for multicast/broadcast services. Satellite Internet service providers already consider including VoIP as one of their service offerings. Therefore, satellite networks will, beyond any doubt, be called to support VoIP, and it appears that they have this ability.

Acknowledgement

This work was supported by the IST FP6 VIVALDI project "Advancing interactive Broadband satellite access by optimal convergence of session based services over DVB-RCS" (FP6-2004-IST-4).

References

[1] ITU-R Recommendation G.114, General Characteristics of International Telephone Connections and International Telephone Circuits: One-way Transmission Time, February 1996.
[2] ETSI TIPHON, End-to-End Quality of Service in TIPHON Systems; Part 2: Definition of Quality of Service (QoS) Classes, TS 101 329-2, July 2000.
[3] M.J. Karam, F.A. Tobagi, Analysis of the delay and jitter of voice traffic over the internet, in: Proceedings of the IEEE INFOCOM 2001, vol. 1, Anchorage, AK, USA, 2001, pp. 824–833.
[4] ETSI DTR/TIPHON-05001, Telecommunications and Internet Protocol Harmonization Over Networks (TIPHON); General Aspects of Quality of Service (QoS), TR 101 329 Ver. 1.2.5, October 1998.
[5] L. Zhang, L. Zheng, K. Ngee, Effect of delay and delay jitter on voice/video over IP, Comput. Commun. 25 (9) (2002) 863–873.
[6] L. Zheng, L. Zhang, D. Xu, Characteristics of network delay and delay jitter and its effect over IP (VoIP), in: Proceedings of the IEEE International Conference on Communications (ICC) 2001, vol. 1, Helsinki, Finland, 2001, pp. 122–126.
[7] R.J.B. Reynolds, A.W. Rix, Quality VoIP – an engineering challenge, BT Technol. J. 19 (2) (2001) 23–32.
[8] ITU-T Contribution D. 110, 1999, Subjective Results on Impairment Effects of Packet Loss, September 1999.
[9] D. Houck, G. Meempat, Call admission control and load balancing for voice over IP, Perform. Eval. 47 (4) (2002) 243–253.
[10] H.T. Tran, T. Ziegler, F. Ricciato, QoS provisioning for VoIP traffic by deploying admission control, Lecture Notes in Computer Science, vol. 2698, Springer Publishing, 2003, pp. 139–153.

[11] T. Uhl, Quality of service in VoIP communication, Int. J. Electron. Commun. 58 (3) (2004) 178–182.
[12] S.G. Wilson, Digital Modulation and Coding, Prentice Hall, NJ, 1996.
[13] T.K. Moon, Error Correction Coding: Mathematical Methods and Algorithms, Wiley–Interscience, 2005.
[14] E. Altman, C. Barakat, Queuing analysis of simple FEC schemes for voice over IP, Comput. Network 39 (5) (2002) 185–206.
[15] J.-C. Bolot, S. Fosse-Parisis, D. Towsley, Adaptive FEC-based error control for internet telephony, in: Proceedings of the IEEE INFOCOM 1999, vol. 3, New York, NY, USA, 1999, pp. 1453–1460.
[16] C. Padhye, K.J. Christensen, W. Moreno, A new adaptive FEC loss control algorithm for voice over IP applications, in: Proceedings of the IEEE performance, computing and communications conference (IPCCC) 2000, Phoenix, AZ, USA, 2000, pp. 307–313.
[17] B. Cheetham, K.M. Nasr, Error concealment for voice over WLAN in converged enterprise networks, in: Proceedings of the IST Mobile Summit 2006, Mykonos, Greece, 2006, pp. 307–313.
[18] C. Perkins, O. Hodson, V. Hardman, A survey of packet loss recovery techniques for streaming audio, IEEE Network 12 (5) (1998) 40–48.
[19] ITU-T Recommendation P.800, Methods for Subjective Determination of Transmission Quality, August 1996.
[20] A. Takahashi, H. Yoshino, N. Kitawaki, Perceptual QoS assessment technologies for VoIP, IEEE Commun. Mag. 42 (7) (2004) 28–34.
[21] ITU-T Recommendation P.861, Objective Quality Measurement of Telephone Band (300–3400 Hz) speech codecs (February 1998).
[22] A.W. Rix, M.P. Holier, The perceptual analysis measurement system for robust end-to-end speech quality assessment, in: Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP) 2000, vol. 3, Istanbul, Turkey, 2000, pp. 1515–1518.
[23] ITU-T Recommendation P.862, Perceptual Evaluation of Speech Quality (PESQ), An Objective Method for End-to-End Speech Quality Assessment of Narrowband Telephone Networks and Speech Codecs, February 2001.
[24] ITU-T Recommendation G.107, The E-Model, A Computational Model for use in Transmission Planning, September 1998.
[25] M.T. Gardner, V.S. Frost, D.W. Petr, Using optimization to achieve efficient quality of service in voice over IP networks, in: Proceedings of the IEEE International Conference on Performance, Computing and Communications Conference (IPCCC) 2003, Phoenix, AZ, USA, 2003, pp. 475–480.
[26] ITU-T Recommendation P.563, Single-ended Method for Objective Speech Quality Assessment in Narrow-band Telephony Applications, May 2004.
[27] ITU-T Recommendation G.711, Pulse Code Modulation (PCM) of Voice Frequencies, November 1988.
[28] ITU-T Recommendation G.723.1, Dual Rate Speech Coder for Multimedia Communications Transmitting at 5.3 and 6.3 kbit/s, March 1996.
[29] ITU-T Recommendation G.723, Extensions of Recommendation G.721 Adaptive Differential Pulse Code Modulation to 24 and 40 kbit/s for Digital Circuit Multiplication Equipment Application, November 1988.
[30] ITU-T Recommendation G.726, 40, 32, 24, 16 kbit/s Adaptive Differential Pulse Code Modulation (ADPCM), December 1990.
[31] ITU-T Recommendation G.728, Coding of Speech at 16 kbit/s using Low-delay Code Excited Linear Prediction, September 1992.
[32] ITU-T Recommendation G.729, Coding of Speech at 8 kbit/s using Conjugate-structure Algebraic-Code-Excited Linear Prediction (CS-ACELP), March 1996.
[33] ITU-T Recommendation G.729A, Reduced Complexity 8 kbit/s CS-ACELP Speech Codec, November 1996.
[34] ITU-T Recommendation G.729B, A Silence Compression Scheme for G.729 Optimized for Terminals Conforming to Recommendation V.70, October 1996.
[35] ITU-T Recommendation G.729D, 6.4 kbit/s CS-ACELP Apeech Coding Algorithm, September 1998.
[36] ITU-T Recommendation G.729E, 11.8 kbit/s CS-ACELP Apeech Coding Algorithm, September 1998.
[37] ETSI, EN 300 961 v.8.0.2, Digital Cellular Telecommunications System (Phase 2+) (GSM); Full Rate Speech; Transcoding, November 2000.
[38] ETSI, EN 300 969 v.8.0.1, Digital Cellular Telecommunications System (Phase 2+) Half Rate Speech; Half Rate Speech Transcoding, November 2000.
[39] ETSI, EN 300 726 v.8.0.1, Digital Cellular Telecommunications System (Phase 2+) (GSM); Enhanced Full Rate (EFR) Speech transcoding, November 2000.
[40] 3GPP TS 26.071, AMR Speech Codec; General Description.
[41] S. Andersen, A. Duric, H. Astrom, R. Hagen, W. Kleijn, J. Linden, Internet Low Bit Rate Codec (iLBC), IETF RFC 3951, December 2004.
[42] iLBC Official Website. <http://www.ilbcfreeware.org>.
[43] ITU-T Recommendation G.722, 7 kHz Audio-Coding within 64 kbit/s, November 1988.
[44] ITU-T Recommendation G.722.1, Low-complexity Coding at 24 and 32 kbit/s for Hands-free Operation in Systems with Low Frame Loss, May 2005.
[45] ITU-T Recommendation G.722.2, Wideband Coding of Speech at Around 16 kbit/s using Adaptive Multi-Rate Wideband (AMR-WB), January 2002.
[46] 3GPP TS 26.273, ANSI-C Code for the Fixed-point Extended Adaptive Multi-Rate-Wideband (AMR-WB+) Speech Codec.
[47] ITU-T Recommendation G.729.1, G.729 based Embedded Variable Bit-rate Coder: An 8–32 kbit/s Scalable Wideband Coder Bitstream Interoperable with G.729, May 1996.
[48] Speex Official Website. <http://www.speex.org>.
[49] J.-M. Valin, The Speex Codec Manual 1.2. <http://www.speex.org/docs/manual/speex-manual/>.
[50] J.-H. Chen, W. Lee, J. Thyssen, RTP Payload Format for BroadVoice Speech Codecs, Internet Draft (<draft-ietf-avt-rtp-bv-04.txt>) (04.04.2005).
[51] J.-H. Chen, J. Thyssen, The broadvoice speech coding algorithm, in: Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing 2007 (ICASSP 2007), vol. 4, Honolulu, HI, USA, 2007, pp. IV-537–IV-540.
[52] B. Goode, Voice over internet protocol (VoIP), Proc. IEEE 90 (9) (2002) 1495–1517.
[53] J. Davidson, J. Peters, M. Bhatia, S. Kalidindi, S. Mukherjee, Voice Over IP Fundamentals, second ed., Cisco Press, 2006.
[54] Sipro Lab Telecom Inc., Pricing Schedule for the Rights of the g.729 Consortium, January 2006.<http://www.sipro.com/pricelist.php>.
[55] H. Schulzrinne, S. Casner, R. Frederic, V. Jacobson, RTP: A Transport Protocol for Real-Time Applications, IETF RFC 1889, January 1996.
[56] H. Oouchi, T. Takenaga, H. Sugawara, M. Masugi, Study on appropriate voice data length of IP packets for VoIP network adjustment, in: Proceedings of the IEEE Global Telecommunications Conference (GLOBECOM) 2002, vol. 2, Taipei, Taiwan, 2002, pp. 1618–1622.
[57] V. Suryavanshi, A. Nosratinia, R. Vedentham, Resilient packet header compression through coding, in: Proceedings of the IEEE Global Telecommunications Conference (GLOBECOM) 2004, vol. 3, Dallas, TX, USA, 2004, pp. 1635–1639.
[58] V. Suryavanshi, A. Nosratinia, Convolutional coding for resilient packet header compression, in: Proceedings of the IEEE Global Telecommunications Conference (GLOBECOM) 2005, vol. 2, St. Louis, MO, USA, 2005, pp. 771–775.
[59] J.A. Ishac, Survey of Header Compression Techniques, NASA/TM-2001-211154, September 2001. <http://gltrs.grc.nasa.gov/cgi-bin/GLTRS/browse.pl?2001/TM-2001-211154.html>.
[60] M. Degermark, B. Nordgren, IP Header Compression, IETF RFC 2507, February 1999.
[61] V. Jacobson, Compressing TCP/IP Headers for Low-Speed Serial Links, IETF RFC 1144, February 1990.
[62] S. Casner, V. Jacobson, Compressing IP/UDP/RTP Headers for Low-Speed Serial Links, IETF RFC 2508, February 1999.
[63] M. Degermark, H. Hannu, K. Svanbro, Evaluation of CRTP performance over cellular radio links, IEEE Personal Commun. 7 (4) (2000) 20–25.
[64] H. Jin, R. Hsu, J. Wang, Performance comparison of header compression schemes for RTP/UPD/IP packets, in: Proceedings of the IEEE Wireless Communications and Networking Conference (WCNC) 2004, vol. 3, Atlanta, GA, USA, 2004, pp. 1691–1696.
[65] G. Mamais, M. Markaki, M.H. Sherif, G. Stassinopoulos, Evaluation of the Casner–Jacobson algorithm for compressing the RTP/UDP/IP headers, in: Proceedings of the Third IEEE Symposium on Computers and Communications (ISCC) 1998, Athens, Greece, 1998, pp. 543–548.
[66] T. Koren, S. Casner, J. Geevarghese, B. Thompson, P. Ruddy, Enhanced Compressed RTP (CRTP) for Links with High Delay, Packet Loss and Reordering, IETF RFC 3545, July 2003.
[67] L.-E. Jonsson, M. Degermark, H. Hannu, K. Svanbro, RObust Checksum-based Header COmpression (ROCCO), Internet Draft (<draft_ietf_rohc_rtp_rocco_01.txt>), June 2000.
[68] Z. Liu, A. Martensson, A. Miyazaki, K. Svanbro, T. Wiebke, T. Yoshimura, H. Zheng, RObust Header Compression (ROHC): Framework and Four Profiles: RTP, UDP, ESP, and Uncompressed, IETF RFC 3095, July 2001.

[69] S. Rein, F.H.P. Fitzek, M. Reisslein, Voice quality evaluation in wireless packet communication systems: a tutorial and performance results for ROHC, IEEE Wireless Commun. 12 (1) (2005) 60–67.
[70] F.H.P. Fitzek, S. Rein, P. Seeling, M. Reisslein, Robust header compression (ROHC) performance for multimedia transmission over 3G/4G wireless networks, Wireless Personal Commun. 32 (1) (2005) 23–41.
[71] L.-E. Jonsson, G. Pelletier, RObust Header Compression (ROHC): A Link-Layer Assisted Profile for IP/UDP/RTP, IETF RFC 3242, April 2002.
[72] Z. Liu, K. Le, Zero-byte Support for Bidirectional Reliable Mode (R-mode) in Extended Link-Layer Assisted RObust Header Compression (ROHC) Profile, IETF RFC 3408, December 2002.
[73] K. Svanbro, J. Wiorek, B. Olin, Voice-over-IP-over-wireless, in: Proceedings of the IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC) 2000, vol. 1, London, UK, 2000, pp. 23–28.
[74] A. Cellatoglu, S. Fabri, S. Worrall, A. Sadka, A. Kondoz, Robust header compression for real-time services in cellular networks, in: Proceedings of the Second International Conference on 3G Mobile Communication Technologies, London, UK, 2001, pp. 124–128.
[75] ITU-T Recommendation H.323, Packet-based Multimedia Communications Systems, June 2006.
[76] J. Glasman, W. Kellerer, H. Müller, Service architectures in H.323 and SIP: a comparison, IEEE Commun. Surv. Tutorials 5 (2) (2003) 32–47. <http://www.comsoc.org/pubs/surveys>.
[77] J. Sinclair, P. Fong, S.M. Harris, M. Walshaw, Configuring Cisco Voice over IP, second ed., Syngress Publishing, 2002.
[78] J. Rosenberg, H. Schulzrinne, G. Camarillo, A. Johnston, J. Peterson, R. Sparks, M. Handley, E. Schooler, SIP: Session Initiation Protocol, IETF RFC 3261, June 2002.
[79] M. Spencer, B. Capouch, E. Guy, F. Miller, K. Shumard, IAX2: Inter-Asterisk eXchange Version 2, Internet Draft (<draft-guy-iax-02.txt>), October 2006.
[80] The Official Advanced Encryption Standard (AES), Federal Information Processing Standards Publication 197, November 2001. <http://www.csrc.nist.gov/publications/fips/fips197/fips>.
[81] F. Andreasen, B. Foster, Media Gateway Control Protocol (MGCP) Version 1.0, IETF RFC 3435, January 2003.
[82] C. Groves, M. Pantaleo, T. Anderson, T. Taylor, Gateway Control Protocol Version 1, IETF RFC 3525, June 2003.
[83] ITU Recommendation H.248.1, Gateway Control Protocol: Version 3, September 2005.
[84] S.A. Baset, H. Schulzrinne, An Analysis of the Skype Peer-to-Peer Internet Telephony Protocol, Technical Report Columbia University, September 2004. <http://www.cs.columbia.edu/techreports/cucs-039-04.pdf>.
[85] H. Schulzrinne, J. Rosenberg, A comparison of SIP and H.323 for internet telephony, in: Proceedings of the Network and Operating System Support for Digital Audio and Video (NOSSDAV), Cambridge, UK, 1998.
[86] B.S. De, P.P. Joshi, V. Sahdev, D. Callahan, End-to-end voice over IP testing and the effects of QoS on signaling, in: Proceedings of the 35th Southeastern Symposium on System Theory (SSST2003), Morgantown, WV, USA, 2003, pp. 142–147.
[87] H. Agrawal, SIP-H.323 Interworking Requirements, IETF Draft, July 2000.
[88] A. Stephens, O.J. Cordell, SIP and H.323 – interworking VoIP networks, BT Technol. J. 19 (2) (2001) 119–127.
[89] H. Liu, P. Mouchtaris, Voice over IP signaling: H.323 and beyond, IEEE Commun. Mag. 38 (10) (2000) 142–148.
[90] A.P. Markopoulou, F.A. Tobagi, M.J. Karam, Assessment of VoIP quality over internet backbones, in: Proceedings of the INFOCOM 2002, vol. 1, New York, USA, 2002, pp. 150–159.
[91] W. Jiang, H. Schulzrinne, Assessment of VoIP availability in the current internet, in: Proceedings of the Workshop on Passive and Active Measurements (PAM2003), La Jolla, CA, USA, 2003.
[92] R. Braden, D. Clark, S. Shenker, Integrated Services in the Internet Architecture: An Overview, IETF RFC 1633, June 1994.
[93] S. Blake, D. Black, M. Carlson, E. Davies, Z. Wang, W. Weiss, An Architecture for Differentiated Services, IETF RFC 2475, December 1998.
[94] F. Wang, P. Mohapatra, Using differentiated services to support internet telephony, Comput. Commun. 24 (18) (2001) 1846–1854.
[95] E. Crawley, R. Nair, B. Rajagopalan, H. Sandlick, A Framework for QoS-based Routing in the Internet, IETF RFC 2386, August 1998.
[96] P. Paul, S.V. Raghavan, Survey of QoS routing, in: Proceedings of the IEEE 15th International Conference on Computer Communication, 2002, Mumbai, Maharashtra, India, 2002, pp. 50–75.
[97] P. Khadivi, S. Samavi, T.D. Todd, H. Saidi, Multi-constraint QoS routing using a new single mixed metric, in: Proceedings of the 2004 IEEE International Conference on Communication (ICC 2004), vol. 4, Paris, France, 2004, pp. 2042–2046.
[98] Z. Wang, J. Crowcroft, Quality-of-service routing for supporting multimedia applications, IEEE J. Select. Areas Commun. 14 (7) (1996) 1228–1234.
[99] M. Curado, E. Monteiro, A survey of QoS routing algorithms, in: Proceedings of the International Conference on Information Technology (ICIT 2004), Istanbul, Turkey, 2004.
[100] A.F. Khalifeh, A. El-Mousa, QoS routing of VoIP using a modified widest-shortest routing algorithm, in: Proceedings of the IEEE/ACS International Conference on Computer Systems and Applications 2007 (AICCSA'07), Amman, Jordan, 2007, pp. 116–123.
[101] L. Geogriadis, R. Guerin, A. Parekh, Optimal multiplexing on a single link: delay and buffer requirements, IEEE Trans. Inform. Theor. 43 (5) (1997) 1518–1535.
[102] M. Katevenis, S. Sidiropoulos, C. Courcoubetis, Weighted round-robin cell multiplexing in a general purpose ATM switch chip, IEEE J. Select. Areas Commun. SAC-9(8) (1991) 1265–1279.
[103] A. Demers, S. Keshav, S. Shenkar, Analysis and simulation of a fair queueing algorithm, in: Proceedings of the ACM SIGCOMM, Austin, TX, USA, 1989, pp. 1–12.
[104] H. Uzunalioglu, D.J. Houck, Y.T. Wang, Call admission control for voice over IP, Int. J. Commun. Syst. 19 (4) (2006) 363–380.
[105] J. Wroclawski, The Use of RSVP with IETF Integrated Services, IETF RFC 2210, September 1997.
[106] S. Wang, Z. Mai, W. Magnussen, D. Xuan, W. Zhao, Implementation of QoS-provisioning system for voice over IP, in: Proceedings of the IEEE Real-time Technology and Application Symposium 2002 (RTAS 2002), San Jose, CA, USA, 2002.
[107] S. Wang, Z. Mai, W. Magnussen, D. Xuan, W. Zhao, Design and implementation of QoS-provisioning system for voice over IP, IEEE Trans. Parallel Distributed Syst. 17 (3) (2006) 276–288.
[108] D. Xuan, C. Li, R. Bettati, J. Chen, W. Zhao, Utilization-based admission control for real-time applications, in: Proceedings of the IEEE International Conference on Parallel Processing 2000, Toronto, Canada, 2000, pp. 251–260.
[109] B.-K. Choi, D. Xuan, R. Bettati, W. Zhao, C. Li, Utilization-based admission control for scalable real-time communication, Real-Time Syst. 24 (2) (2002) 171–202.
[110] B.T. Doshi, D. Eggenschwiler, A. Rao, B. Samadi, Y.T. Wang, J. Wolfson, VoIP network architectures and QoS strategy, Bell Labs Tech. J. 7 (4) (2003) 41–59.
[111] A. Farrel, A. Ayyangar, J. Vasseur, Inter-Domain MPLS and GMPLS Traffic Engineering – Resource Reservation Protocol-Traffic Engineering (RSVP-TE) Extensions, IETF RFC 2210, February 2008.
[112] B. Jamoussi, L. Andersson, R. Callon, R. Dantu, L. Wu, P. Doolan, T. Worster, N. Feldman, A. Fredette, M. Girish, E. Gray, J. Heinanen, T. Kilty, A. Malis, Constraint-Based LSP Setup using LDP, IETF RFC 3212, January 2002.
[113] Z. Zhang, Z. Duan, L. Gao, Y. Hou, Decoupling QoS control from core routers: a novel bandwidth broker architecture for scalable support of guaranteed services, in: Proceedings of the ACM SIGCOMM 2000, Stockholm, Sweden, 2000, pp. 71–83.
[114] L. Breslau, S. Jamin, S. Shenker, Comments on the performance of measurement-based admission control algorithms, in: Proceedings of the IEEE INFOCOM 2000, vol. 3, Tel-Aviv, Israel, 2000, pp. 1233–1242.
[115] K. Shiomoto, N. Yamanaka, T. Takahashi, Overview of measurement-based connection admission control methods in ATM networks, IEEE Commun. Surv. Tutorials 2 (1) (2002) 2–13. <http://www.comsoc.org/livepubs/surveys>.
[116] K. Mase, Y. Toyama, A.A. Bilhaj, Y. Suda, QoS management for VoIP networks with edge-to-edge admission control, in: Proceedings of the IEEE Globecom 2001, vol. 4, San Antonio, USA, 2001, pp. 2556–2560.
[117] K. Mase, Y. Toyama, End-to-end measurement based admission control for VoIP networks, in: Proceedings of the IEEE International Conference on COmmunications ICC 2002, vol. 2, New York, USA, 2002, pp. 1194–1198.
[118] V. Elek, G. Karlsson, R. Ronngren, Admission control based on end-to-end measurements, in: Proceedings of the IEEE INFOCOM 2000, vol. 3, Tel-Aviv, Israel, 2000, pp. 1233–1242.
[119] K. Mase, Towards scalable admission control for VoIP networks, IEEE Commun. Mag. 42 (7) (2004) 42–47.

[120] D.J. Houck, E. Kim, H. Uzanalioglu, L.A. Wehr, A measurement-based admission control algorithm for VoIP, Bell Labs Tech. J. 8 (2) (2003) 77–110.
[121] D.R. Jeske, B. Samadi, K. Sohraby, Y.-T. Wang, Q. Zhang, QoS with edge-based call admission control in IP networks, Lecture Notes in Computer Science, vol. 2345, Springer Publishing, 2002, pp. 178–189.
[122] G. Bianchi, F. Borgonovo, A. Ce, C. Petrioli, Endpoint admission control with delay variation measurements for QoS in IP networks, ACM SIGCOMM Computer Communications Review 32 (2) (2002) 61–69.
[123] K. Burst, L. Joiner, G. Grimes, Delay based congestion detection and admission control for voice quality in enterprise or carrier controlled IP networks, IEEE eTrans. Network Service Manage. 2 (1) (2005) 1–8.
[124] D. Houck, H. Uzunalioglu, VoIP: link-based management in MPLS networks, in: Proceedings of the MPLS World Congress, Paris, France, 2005.
[125] H. Uzunalioglu, K.P. Das, D.J. Houck, An architecture and admission control algorithm for multi-precedence voice over IP (VoIP) calls, in: Proceedings of the IEEE Military Communications Conference 2004 (MILCOM 2004), Monterey, CA, USA, 2004, pp. 906–912.
[126] S. Jamin, P.B. Danzig, S. Shenker, L. Zhang, A measurement-based admission control algorithm for integrated services packet networks, in: Proceedings of the ACM SIGCOMM'95, Cambridge, MA, USA, 1995, pp. 2–13.
[127] S. Jamin, P.B. Danzig, S. Shenker, L. Zhang, A measurement-based admission control algorithm for integrated services packet networks, IEEE/ACM Trans. Network. 5 (1) (1997) 56–70.
[128] R.J. Gibbens, F.P. Kelly, Distributed connection acceptance control for a connectionless network, in: Proceedings of the 16th International Teletraffic Congress, Edinburgh, Scotland, 1999, pp. 941–952.
[129] A. Jamalipour, J. Kim, Measurement-based admission control scheme with priority and service classes for application in wireless IP networks, Int. J. Commun. Syst. 16 (6) (2003) 535–551.
[130] Y.-W. Leung, Dynamic bandwidth allocation for internet telephony, Comput. Commun. 29 (18) (2006) 3710–3717.
[131] T. Porter, J. Kanclirz, A. Zmolek, A. Rosela, M. Cross, L. Chaffin, B. Baskin, C. Shim, Practical VoIP Security, Syngress Publishing, 2006.
[132] R. Stanton, Secure VoIP – an achievable goal, Comput. Fraud Security 2006 (4) (2006) 11–14.
[133] D. Sisalem, J. Kuthan, S. Ehlert, Denial of service attacks targeting a SIP VoIP infrastructure: attack scenarios and prevention mechanisms, IEEE Network 20 (5) (2006) 26–31.
[134] D. Shin, J. Ahn, C. Shim, Progressive multi gray-leveling: a voice spam protection algorithm, IEEE Network 20 (5) (2006) 18–24.
[135] N. Thanthry, R. Pendse, K. Namuduri, Voice over IP security and law enforcement, in: Proceedings of the 39th Annual 2005 International Carnahan Conference on Security Technology (ICCST' 05), Las Palmas de Gran Canaria, Spain, 2005, pp. 246–250.
[136] F. Cao, S. Malik, Security analysis and solutions for deploying IP telephony in the critical infrastructure, in: Proceedings of the First International Conference on Security and Privacy for Emerging Areas in Communication Networks 2005 (SecureComm 2005), Athens, Greece, 2005, pp. 171–180.
[137] F. Cao, S. Malik, Vulnerability analysis and best practices for adopting IP telephony in critical infrastructure sectors, IEEE Commun. Mag. 44 (4) (2006) 138–145.
[138] D. McGrew, M. Naslund, E. Carrara, K. Norrman, The Secure Real-time Transport Protocol (SRTP), IETF RFC 3711, March 2004.
[139] Z. Anwar, W. Yurcik, R.E. Johnson, M. Hafiz, R.H. Campbell, Multiple design patterns for voice over IP (VoIP) security, in: Proceedings of the 25th IEEE International Performance, Computing, and Communications Conference 2006 (IPCCC 2006), Phoenix, AZ, USA, 2006, pp. 485–492.
[140] H. Abdelnur, V. Cridlig, R. State, O. Festor, VoIP security assessment: methods and tools, in: Proceedings of the First IEEE Workshop on VoIP Management and Security 2006 (VoIP MaSe 2006), Vancouver, Canada, 2006, pp. 29–34.
[141] I.S. 802.11-1997, Information Technology – Telecommunications and Information Exchange Between Systems-Local and Metropolitan Area Networks-specific Requirements-Part 11: Wireless Lan Medium Access Control (MAC) and Physical Layer (PHY) Specifications, November 1997.
[142] S.R. Fluhrer, I. Mantin, A. Shamir, Weaknesses in the key scheduling algorithm of RC4, in: Proceedings of the Eighth Annual Workshop on Selected Areas in Cryptography 2001 (SAC 2001), Toronto, Ontario, Canada, 2001, pp. 1–24.
[143] I.802.1X-2004, IEEE Standards for Local and Metropolitan Area Networks: Port Based Network Access Control, December 2004.
[144] B. Aboba, L. Blunk, J. Vollbrecht, J. Carlson, H. Levkowetz, Extensible Authentication Protocol (EAP), IETF RFC 3748, June 2004.
[145] I.P802.11i/D10.0, Medium Access Control (MAC) Security Enhancements, Amendment 6 to IEEE Standard for Information Technology Telecommunications and Information Exchange Between Systems Local and Metropolitan Area Networks Specific Requirements Part 11: Wireless Medium Access Control (MAC) and Physical Layer (PHY) Specifications, April 2004.
[146] R. Barbieri, D. Bruschi, E. Rosti, Voice over IPsec: analysis and solutions, in: Proceedings of the 18th Annual Computer Security Applications Conference (ACSAC'02), Las Vegas, USA, 2002, pp. 261–270.
[147] A. Nascimento, A. Passito, E. Mota, Can i add a secure VoIP call? in: Proceedings of the International Symposium on a World of Wireless, Mobile and Multimedia Networks, 2006 (WoWMoM 2006), Monterey, CA, USA, 2006.
[148] W.C. Barker, Recommendation for the Triple Data Encryption Algorithm (TDEA) Block Cipher, NIST (National Institute of Standards and Technology) Special Publication 800-67, May 2004. <http://csrc.nist.gov/publications/nistpubs/800-67/SP800>.
[149] E.T. Aire, B.T. Maharaj, L.P. Linde, Implementation considerations in a SIP based secure voice over IP network, in: Proceedings of the Seventh AFRICON Conference in Africa (AFRICON 2004), vol. 1, Gaborone, Botswana, 2004, pp. 167–172.
[150] J. Fiedler, T. Kupka, S. Ehlert, T. Magedanz, D. Sisalem, VoIP defender: highly scalable SIP-based security architecture, in: Proceedings of the First International Conference on Principles, Systems and Applications of IP Telecommunications (IPTComm 2007), New York, USA, 2007.
[151] S. Ehlert, G. Zhang, T. Magedanz, Increasing SIP firewall performance by ruleset size limitation, in: Proceedings of the Workshop on VoIP Technology: Research and Standards for Reliable Applications, Cannes – Palais de Festivals, France, 2008.
[152] S. Spinsante, E. Gambi, E. Bottegoni, Security solutions in VoIP applications: state of the art and impact on quality, in: Proceedings of the IEEE International Symposium on Consumer Electronics 2008 (ISCE 2008), Algarve, Portugal, 2008, pp. 1–4.
[153] N. Sulaiman, R. Carrasco, G. Chester, Impact of security on voice quality in 3G networks, in: Proceedings of the Third IEEE Conference on Industrial Electronics and Applications 2008 (ICIEA 2008), Singapore, 2008, pp. 1583–1587.
[154] H. Xiao, P. Zarrella, Quality effects of wireless VoIP using security solutions, in: Proceedings of the IEEE Military Communications Conference 2004 (MILCOM 2004), vol. 3, Monterey, CA, USA, 2004, pp. 1352–1357.
[155] M. Boulmalf, E. Barka, A. Lakas, Analysis of the effect of security on data and voice traffic in WLAN, Comput. Commun. 20 (11–12) (2007) 2468–2477.
[156] M.A.V. Castro, F.J.G. Serrano, A.M. Fernández, G.M. Moya, Quality of service of VoIP over DVB-RCS, in: Proceedings of the Sixth Baiona Workshop on Signal Processing in Communications, vol. 2, Baiona, Spain, 2003.
[157] J. Janssen, D.D. Vleeschauwer, G.H. Petit, R. Windey, J.-M. Leroy, Delay bounds for voice over IP calls transported over satellite access networks, Mobile Network Appl. 7 (2) (2002) 79–89.
[158] J. Ott, S. Prelle, D. Meyer, O. Bergmann, IPTEL-via-Sat: Cookbook for IP Telephony via DVB-RCS, SatLabs Group, May 2005. <http://satlabs.org/documents>.
[159] A. Jun, J. Prat, AMERHIS: DVB-RCS meets Mesh Connectivity, SatLabs Group. <http://satlabs.org/documents>.
[160] T. Nguyen, F. Yegenoglou, A. Sciuto, R. Subbarayan, Voice over IP service and performance in satellite networks, IEEE Commun. Mag. 39 (3) (2001) 164–171.
[161] A. Vermesan, H.P. Lexow, H. Skinnemoen, J. Hetland, VoIP Over DVB-RCS: A Radio Resource and QoS Perspective, SatLabs Group (Voip White Paper by Nera Broadband Satellite AS), December 2004. <http://satlabs.org/documents>.
[162] M. Lambert, VoIP Over Satellite, SatLabs Group (VoIP TEchnical Note by EMS Satellite Networks), May 2004. <http://satlabs.org/documents>.
[163] D. Sisalem, M. Corici, S. Ehlert, R. Popescu-Zeletin, VDSat: Nomadic Satellite-Based VoIP Infrastructure, in: Proceedings of the Second International Symposium on Wireless Communication Systems, 2005, Siena, Italy, 2005, pp. 619–623.
[164] V.Y.H. Kueh, R. Tafazolli, B. Evans, Enhancing the radio link protocol for VoIP session establishment signalling over satellite-

UMTS, in: Proceedings of the IEEE 59th Vehicular Technology Conference, 2004 (VTC 2004-Spring), vol. 5, Milan, Italy, 2004, pp. 2787–2791.
[165] T.E.B. deMello, A. Kiefer, Satellite VoIP access gateway, in: Proceedings of the International Conference on Digital Telecommunications 2006 (ICDT'06), Côte d' Azur, France, 2006.
[166] H. Cruickshank, A. Sánchez, Z. Sun, B. Carro, Voice over IP over satellite links, in: Proceedings of the Eighth IEEE International Conference on Electronics, Circuits and Systems 2001 (ICECS 2001), vol. 1, Malta, 2005, pp. 473–476.
[167] R. Toegl, U. Birnbacher, O. Koudelka, Deploying IP telephony over satellite links, in: Proceedings of the Second International Symposium on Wireless Communication Systems 2005, Siena, Italy, 2005, pp. 624–628.
[168] R. Deininger, K. Griep, D. Morse, S. Million, S. Knapp, Investigation of party line voice over INMARSAT's mobile packet data service, in: Proceedings of the 23rd Digital Avionics Systems Conference 2004 (DASC 04), vol. 1, Salt Lake City, USA, 2004, pp. 1–10.
[169] G. Giambene (Ed.), Resource Management in Satellite Networks, Springer, 2007.
[170] H. Skinnemoen, A. Vermesan, A. Iuoras, G. Adams, X. Lobao, VoIP over DVB-RCS with QoS and bandwidth on demand, IEEE Wireless Commun. 12 (5) (2005) 46–53.
[171] H. Skinnemoen, R. Leirvik, J. Hetland, H. Fanebust, V. Paxal, Interactive IP-network via satellite DVB-RCS, IEEE J. Select. Areas Commun. 22 (3) (2004) 508–517.
[172] A.G. Berumen, M. Marot, Cross-layer algorithm for VOIP applications over satellite, in: Proceedings of the IEEE 18th International Symposium on Personal, Indoor and Mobile Radio Communications 2007 (PIMRC 2007), Athens, Greece, 2007.
[173] M. Kalama, G. Acar, B. Evans, A. Isoard, VoIP over DVB-RCS satellite systems: trial results and the impact of adaptive speech coding using cross-layer design, Comput. Network 52 (13) (2008) 2461–2472.
[174] Z. Sun, A. Sánchez, H. Cruickshank, B. Carro, M. Vázquez, VoIP real-time flows in satellite networks with OBP, in: Proceedings of the Seventh International Workshop on Digital Signal Processing Techniques for Space Communications (DSP2000), Sesimbra, Portugal, 2001.
[175] S. Karapantazis, F.-N. Pavlidou, QoS handover management for multimedia LEO satellite systems, Telecommun. Syst. 32 (4) (2006) 225–245.
[176] S. Bayhan, G. Gür, F. Alagöz, VoIP performance in multi-layered satellite IP networks with on-board processing capability, in: Proceedings of the Seventh IEEE International Symposium on Computer Networks (ISCN'06), vol. 1, Istanbul, Turkey, 2006, pp. 197–202.
[177] A. Durresi, D. Dash, B.L. Anderson, R. Kannan, S. Kota, R. Jain, Routing of real-time traffic in a transformational communications architecture, in: Proceedings of the IEEE Aerospace Conference 2004, vol. 2, Istanbul, Turkey, 2003, pp. 1086–1104.
[178] D.S. Dash, A. Durresi, R. Jain, Routing of VoIP traffic in multi-layered satellite networks, in: Proceedings of the SPIE Performance and Control of Next-Generation Communications Networks, ITCOMM 2003, vol. 5344, Orlando, FL, USA, 2003, pp. 65–75.
[179] S. Iyengar, H. Cruickshank, P. Pillai, G. Fairhurst, L. Duquerroy, Security requirements for IP over satellite DVB networks, in: Proceedings of the 16th IST Mobile and Wireless Communications Summit, 2007, vol. 1, Budapest, Hungary, 2007, pp. 1–6.
[180] J. Oetting, K. King, The impact of IPSec on DoD teleport throughput efficiency, in: Proceedings of the IEEE Military Communications Conference 2004 (MILCOM 2004), vol. 2, Monterey, CA, USA, 2004, pp. 7176–721.
[181] L. Duquerroy, S. Josset, O. Alphand, P. Berthou, T. Gayraud, SatIPSec: an optimized solution for securing multicast and unicast satellite transmissions, in: Proceedings of the 22nd AIAA International Communications Satellite Systems Conference 2004 (ICSSC 2004), California, USA, 2004.
[182] L. Fan et al., Next-generation satellite systems with IPv6 and DVB, in: Proceedings of the 25th AIAA International Communications Satellite Systems Conference 2007 (AIAA 2007), Seoul, Korea, 2007.
[183] H. Cruickshank, S. Iyengar, S. Combes, L. Duquerroy, G. Fairhurst, M. Mazzella, Security requirements for IP over satellite DVB networks, in: Proceedings of the 16th IST Mobile and Wireless Communications Summit 2007, Budapest, Hungary, 2007.

Stylianos Karapantazis received a Diploma (in 2003) and a Ph.D. degree (in 2007) in Electrical and Computer Engineering from the Aristotle University of Thessaloniki, where he is currently a researcher. His interests lie in the fields of radio resource management and Multicast/Broadcast protocols for High Altitude Platforms (HAPs), call admission control, handover management and routing in satellite networks, as well as traffic management and header compression for Voice over IP (VoIP) networks. He has served as a reviewer in several journals and conferences. He has been involved in several projects supported by the Greek General Secretariat for Research and Development as well as in European projects in regard to HAP and satellite communications.

Fotini-Niovi Pavlidou received the Ph.D. degree in Telecommunications Networks (1988) and the Diploma in Mechanical/Electrical Engineering (1979) from the Aristotle University of Thessaloniki (AUTh), Greece, where she is currently a Professor engaged in teaching and research. Her interests include traffic analysis and design of wireless networks, performance evaluation and QoS, multimedia applications over the Internet, satellite and high altitude platforms, etc. She has been the Delegate of Greece in the European COST Program on Telecommunications (1998–2006) and served as Chairperson for the COST262 Action "Spread Spectrum systems and techniques for wired and wireless System". She is involved in many European Research and Educational Projects (Telematics, Tempus, ACTS, IST FP5, IST FP6, etc.). She is a Senior Member of IEEE currently chairing the Joint VTS & AES Greece Chapter. She has also served as the Chairperson of the several International conferences. Furthermore, she currently serves as Editor for the JCN (Journal of Communications Networks) and Wireless Personal Communications journals and has served in the editorial board of many special issues.