Choosing the UMTS Air Interface Parameters, the Voice Packet Size and the Dejittering Delay for a Voice-over-IP Call between a UMTS and a PSTN Party

Fabrice Poppe, Danny De Vleeschauwer and Guido H. Petit
Alcatel Bell, Corporate Research Center
Francis Wellesplein 1
B-2018 Antwerp, Belgium
Tel.: +32 3 240 80 06 – Fax: +32 3 240 99 32
e-mail: {fabrice.poppe, danny.de_vleeschauwer, guido.h.petit}@alcatel.be

Abstract--In this paper we develop a methodology to set the VoIP application parameters (voice packet size and dejittering delay) and the UMTS air interface parameters in such a way that the quality of VoIP calls involving a UMTS party is ensured. We use analytical techniques to determine the delay and packet loss contributions of the various transmission stages crossed by the voice packet flow, and the E-model to predict the perceived quality. Our numerical results show that provided the parameters of the VoIP application and the UMTS air interface are chosen properly, UMTS access and the stringent delay and packet loss requirements of VoIP are reconcilable.

Index terms--UMTS, Voice-over-IP, prediction of perceived voice quality, guaranteeing QoS to delay and loss sensitive applications.

I. INTRODUCTION

A. Background

IP-centric network architectures are gradually becoming a reality. An illustration of this is the increasing interest in voice-over-IP (VoIP), which would allow the integration of the separate networks in place today for voice and data transport [12]. The potential benefits for corporate and home users include the reduced cost of only needing a single connection to the outside world as well as lower per-minute telephony rates. For operators it has the benefit of having to manage only one network for both voice and data traffic.

At the same time, standardisation of third-generation mobile communication systems is rapidly progressing. Universal Mobile Telecommunications System (UMTS) standardised by 3GPP enhances the services provided by current second-generation systems such as Global System for Mobile Communications (GSM) with high-rate data services (2 Mbit/s for indoor, 384 kbit/s for outdoor in metropolitan area, and 144 kbit/s for outdoor in rural area). These larger bit rates will enable wireless operators to offer much more services to their customers, such as web browsing, e-mail,… This is an incentive for studying whether voice can be carried over the UMTS air interface in IP packets, to obtain a single solution for data and voice services.

B. Overview of Previous Work and Contributions

Setting up a VoIP call requires a decision on several parameters (such as type of codec, echo control, voice packet size and dejittering delay) [1]. Using the E-model [8] as a tool to predict the perceived quality of the call, it can be shown that voice quality can only be acquired if the echo is controlled and strict bounds on the packet loss and the mouth-to-ear (M2E) delay (the delay incurred by the voice signal from the moment it is uttered by the speaker till the moment it is heard by the listener) are respected [4]. Setting the parameters of the VoIP call thus requires a thorough understanding of the delay and loss mechanisms affecting the voice packet flow between the parties involved in the call.

Two wireline access scenarios were studied in [3]. In the first scenario, both parties are connected via the PSTN network to an IP backbone (PSTN-to-PSTN scenario), with VoIP gateways performing the interworking between the PSTN and the IP backbone. In the second scenario, both parties conduct the VoIP call directly from their PC (PC-to-PC scenario), and are connected to an IP backbone through an access network, e.g. an ADSL access line. For both scenarios an analytical expression for the mouth-to-ear delay is derived. These analytical expressions are then used to set the packet size and the dejittering delay.

Wireless access scenarios were studied in [10] (access through a Low Earth Orbit Satellite (LEOS) network or a Geostationary Earth Orbit Satellite (GEOS) network) and in [13] (GPRS access). The focus in these papers was on delay and its impact on voice quality, not on packet loss. However, packet loss is a challenging problem in wireless access scenarios. The unreliability of the wireless channels forces one to use techniques that introduce additional delay (such as interleaving) or decrease bandwidth efficiency (such as channel coding). In [14] we studied the trade-off of packet loss versus packet delay over the UMTS air interface assuming that Forward Error Correction (FEC) and interleaving are used to mask transitory bad conditions of the wireless channel. In this paper we build on these results to develop a method for setting the voice packet size, the dejittering delay and the UMTS air interface parameters such that the quality of VoIP calls involving a UMTS party is guaranteed from the onset of the call.

C. Overview of the Contents of the Paper

We consider a VoIP call between a UMTS subscriber and a PSTN subscriber. In Section II, we describe which transmission stages the voice packet flow associated with such a call encounters, and list the M2E delay and packet loss contributions of each of these stages. We briefly summarise the provisions that have been made in the design of the UMTS air interface to deal with the unreliability inherent to a wireless access network.

In Section III and Section IV, we study how the subjective voice quality of a VoIP call between a UMTS and a PSTN party can be predicted. In Section III, we briefly review the E-model. We use the E-model to calculate a rating for the VoIP calls. This rating depends on the delay and packet loss contributions of the transmission stages the voice packets go through. In Section IV, we show how the delay and packet loss contribution of each transmission stage can be calculated starting from the parameters describing its operation.

In Section V, we develop a methodology for choosing the UMTS air interface parameters, the dejittering delay and the voice packet size such that the quality of the voice call is guaranteed. We provide detailed numerical results.

Finally we list the conclusions we derive from this study in Section VI.

II. A PACKETISED VOICE CALL BETWEEN A UMTS AND A PSTN PARTY

A. A UMTS-to-PSTN Scenario

We consider a VoIP call between a UMTS subscriber and a PSTN subscriber. We assume that the PSTN subscriber can reach, via the PSTN, a VoIP gateway (GW) that is connected to an IP network. Fig. 1 lists the contributions to the mouth-to-ear delay and the voice packet loss of each of the transmission stages the voice signal crosses. A detailed analysis of how each of these contributions can be calculated will be made in Section IV.

The Mobile Terminal (MT) of the UMTS subscriber encodes voice frames, packetises voice code words into IP packets and transmits these packets over the air interface. Between the MT and the VoIP gateway, the voice call goes over IP, first crossing the air interface, then the UMTS terrestrial radio access network (UTRAN) and finally the IP backbone. An MT accesses the UTRAN over the air interface via a base station (node B). Between node B and the ingress router R to the IP backbone network, IP packets traverse one or more radio network controllers (RNCs). Fig. 2 shows the protocols which come into play between the MT and the VoIP gateway. Macro-diversity [2] (soft hand-overs) requires a fast switching technology in the terrestrial radio access network. Therefore, IP packets will traverse the UTRAN on top of AAL2 and the delay incurred in the UTRAN will be small (5-10 ms).

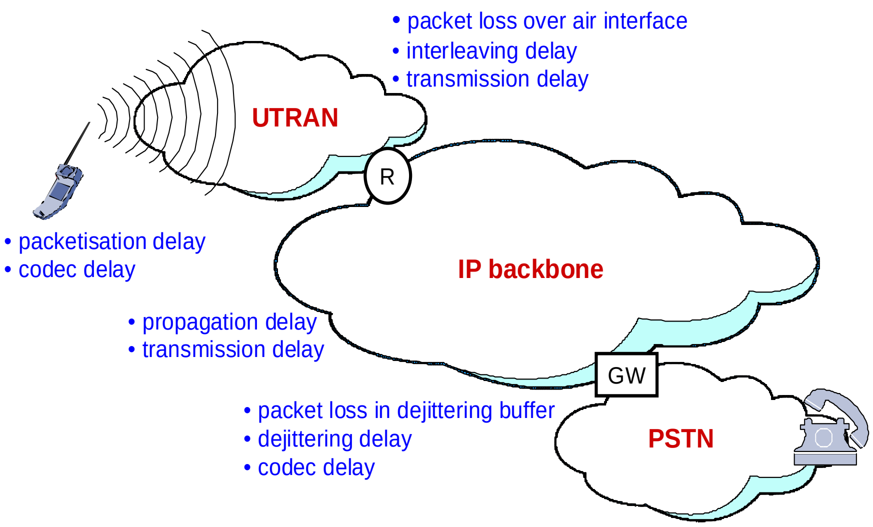

Fig. 1. Mouth-to-ear delay and packet loss contributions of the mobile terminal, the UMTS air interface and terrestrial radio access network, the IP backbone and the VoIP gateway in a UMTS-to-PSTN scenario.

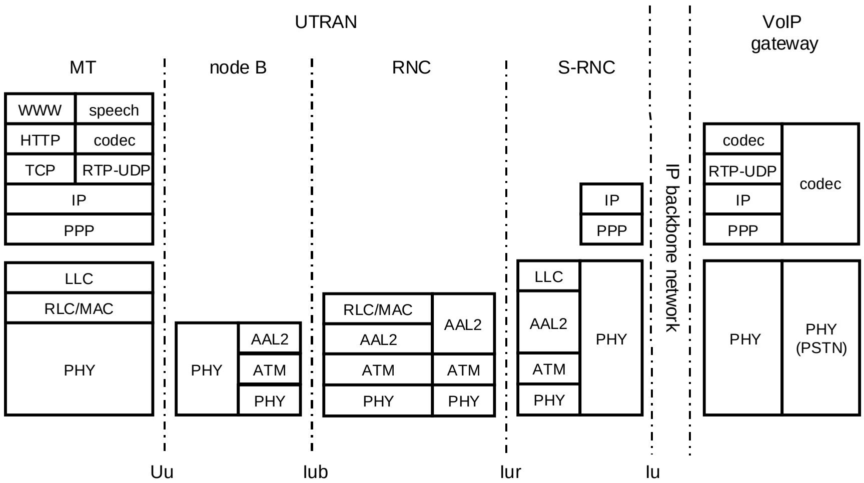

Fig. 2. Protocol stack from the mobile terminal up to the VoIP gateway.

B. The UMTS Air Interface

We consider the wireless uplink (the downlink is similar) and trace the path followed by voice frames, from the MT, through the various protocol layers constituting the air interface, to node B in the UTRAN. The MT encodes voice frames, packetises voice code words into IP packets and transmits these packets over the air interface. In the Radio Link Control (RLC) layer the voice packets are segmented into transport blocks (RLC PDUs). The blocks are handed over to the Medium Access Control (MAC) layer, where they may have to queue before they are handed over to the physical layer. The physical layer handles the transmission of blocks over the physical channel.

Wide-band Code-Division Multiple Access (WCDMA) has been chosen as the basic radio-access technology for UMTS. Information is transmitted over the physical channel (air interface) in radio frames with a duration of 10 ms, each consisting of 16 time slots. In the uplink, every time slot contains 10×2^k bits, with k an integer between 0 and 6 (16-1024 kbit/s). In the downlink, every time slot contains $10\times2^{k+1}$ bits, again with k an integer between 0 and 6 (32-2048 kbit/s).

Two phenomena cause the power of the signal received by/from a MT to vary in time. The slow variations of the instantaneous power level around the area mean power level, because of e.g. man-made structures that temporarily hide the transmitter from the view of the receiver, are called shadowing. The fast variation of the instantaneous signal strength around the local mean power that is caused by multipath reflections by local scatterers, is called fading. Fading and shadowing introduce error bursts rather than isolated bit errors.

The physical layer of the air interface has the difficult task of making the unreliable character of the wireless channel imperceptible for applications that tolerate only small packet losses. The functions performed by the physical layer to this end are listed in [2]. In this section, we briefly review how the settings of the engineering parameters associated with these functions affect the Quality of Service (QoS) applications receive and the bandwidth efficiency over the UMTS air interface.

The service offered by the physical layer of the air interface can be tailored in three ways. The first is the possibility to append a Cyclic Redundancy Check (CRC) to each of the transport blocks. The CRC is meant to be used in conjunction with an Automatic Repeat reQuest (ARQ) mechanism. For a streaming application like VoIP, using CRCs makes little sense, since there is little time for retransmissions.

After CRCs have been appended to individual blocks, the bit stream is channel coded. The channel codes can be chosen from among a number of alternatives with different coding rates (thus different performance and bandwidth consumption). The choice depends on the sensitivity of the application to packet loss.

Finally, the service an application receives also depends on the interleaving scheme. The physical layer applies a two-stage (inter-frame and intra-frame) bit-level interleaving to mitigate the effect of transitory bad channel conditions. Inter-frame interleaving happens on the scale of an interleaving span, also called Transmission Time Interval (TTI). The TTI has a fixed duration for each transport channel, equal to 1, 2, 4 or 8 radio frames (10-80 ms). During each TTI, the blocks handed over by the MAC layer during the previous TTI are transmitted.

The choice of the TTI for an application is a trade-off between packet loss and packet delay. A longer TTI means more delay, but also decreases the probability that an error burst on the physical channel affects the application: successive bits are spread out over a larger time period,

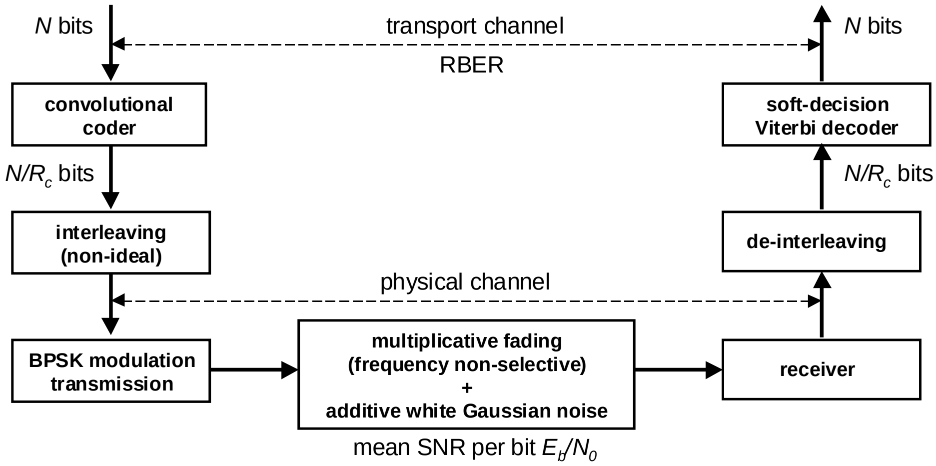

Fig. 3. Model of the transport channel.

which decreases the impact of transitory bad channel conditions. How large the TTI can be made depends on the mouth-to-ear delay budget. How small it can be made depends on the power budget of the MT and the tolerance of the codec towards the loss of voice frames.

The packet loss ratio depends on the Residual Bit Error Rate (RBER) seen on the wireless transport channel. The RBER depends on the channel coding and the interleaving performed by the physical layer, and on the characteristics of the physical channel. In [14] we developed a method to determine RBER, based on a model of the wireless transport channel shown in Fig. 3.

III. The E-model

Given a set of parameters characterising the transmission of the voice packets, the E-model [8] predicts the subjective quality that will be experienced by an "average" listener. Because the E-model is a cornerstone of our methodology for engineering the VoIP application parameters, we summarise here the aspects that are of importance to us. A detailed description can be found in [4].

The E-model is used to calculate a rating for packetised telephony calls, starting from objective parameters describing the operation of the transmission stages the voice packets go through. Every rating value corresponds to a speech transmission category, as shown in Table 1 [8].

The value R of the rating can be calculated as

$$R = R_{intrinsic} + A - I_d(T_{M2E}, EL) - I_e(P_{loss}) . \qquad (1)$$

In this expression, $R_{intrinsic}$ groups the effects of noise and impairments occuring simultaneously with the signal. The advantage A is the deterioration that the callers are willing to tolerate because of a convenience offered to them during the call, for example mobility.

The impairments introduced by delay are grouped in $I_d(T_{M2E},EL)$. This impairment depends on the echo loss EL and the mouth-to-ear delay T_{M2E}. When the echo is not carefully controlled, the delay impairment $I_d(T_{M2E},EL)$ becomes very large, even for the tiniest amount of mouth-to-ear delay. Standardised values for the delay impairment can be found in [7]. Assuming that the distortion impairment is zero, the rating R drops below 70 at a mouth-to-ear delay of 25 ms for an echo loss of EL=21 dB, and at a mouth-to-ear delay of 400 ms for perfect echo control, that is, $EL=\infty$.

The impairments introduced by distortion are brought together in $I_e(P_{loss})$. Assuming there is no voice packet loss, the distortion is equal to the distortion introduced by the encoder (intrinsic impairment of the codec). When there is packet loss, the distortion impairment increases with the packet loss probability P_{loss}. Standardised values for the distortion impairment of a variety of codecs can be found in [4] and [9].

Table 1. Speech transmission quality categories and corresponding rating value ranges.

R-value Range	100-90	90-80	80-70	70-60	60-0
speech transmission quality category	best	high	medium	low	poor

PSTN quality

IV. The Working Point of a VoIP Call between a UMTS and a PSTN Subscriber

With a given codec, a given set of UMTS air interface parameters (convolutional code and interleaving span), a given mobile power budget, a given speed of the mobile party, and a given set of VoIP application parameters (voice packet size and dejittering delay) corresponds a point in the (T_{M2E}, I_e)-plane. We call this point the working point of the VoIP application.

To determine the location and the rating of the working point of the VoIP application, we need to calculate the mouth-to-ear delay T_{M2E}, and the packet loss probability P_{loss}.

A. *Calculation of the Mouth-to-Ear delay*

1) *Coding and packetisation.*

The Mobile Terminal encodes voice frames, packetises voice code words into IP packets and transmits these over the air interface.

The voice codec is characterised by the duration T_W (in ms) of the voice frames it encodes, the size B_W (in bits) of the voice code words it produces and the resulting bitrate $R_{cod}=B_W/T_W$ (in kbit/s), and possibly also by the duration T_{LA} (in ms) of its look-ahead interval. The encoding delay T_{enc} is approximately given by T_W+T_{LA}. Assuming that a packet contains N_W voice code words, the packetisation delay is given by

$$T_{pack} = N_W T_W = \frac{N_W B_W}{R_{cod}} \ . \tag{2}$$

Each voice packet contains $O_{RTP/UDP/IP}$ overhead bits and a number of voice frames. The voice packet size B_P is given by

$$B_P = O_{RTP/UDP/IP} + N_W B_W \ . \tag{3}$$

Assuming that no header compression is performed, the size $O_{RTP/UDP/IP}$ of the RTP/UDP/IP overhead is (12+8+20) $\times$ 8 bits. We assume that PPP is used for the transport of the voice packets. The size B_F of the PPP frames is given by

$$B_F = B_P + O_{PPP} \tag{4}$$

where $O_{PPP}=7 \times 8$ bits is the size of the PPP frame overhead. The filling factor Φ is defined as

$$\Phi = \frac{N_W B_W}{N_W B_W + O_{RTP/UDP/IP} + O_{PPP}} \ . \tag{5}$$

From a bandwidth efficiency point of view, it is beneficial to make the size of the voice packets, that is, N_W, as large as possible.

2) *UMTS air interface and UTRAN.*

We assume that voice packets do not have to queue in the MAC layer of the air interface. This is a valid assumption when the network is aware of the QoS requirements of the VoIP application and provisions capacity at the time it is needed.

We assume that voice packets receive Head-of-Line (HoL) priority at the PPP layer, that fragmentation-oriented PPP is used, and that the maximum transfer unit for data traffic is chosen in such a way that voice packets never have to wait for more than the duration of a radio frame. We call this delay the MT queueing delay $T_{q,MT}$.

Let R_T stand for the capacity of the wireless transport channel, N_I for the number of radio frames in each interleaving span, and T_R=10 ms for the duration of a radio frame. The physical layer of the air interface introduces a minimum delay $T_{m,AI}$ of $N_I T_R$ (see [14]), that is, one TTI. Apart from that it also introduces a serialisation delay, given by B_F/R_T.

Unfortunately, the physical layer of the UMTS air interface also introduces jitter, which has to be compensated in the dejittering buffer of the VoIP gateway. Packets arriving just before the start of a TTI will be served almost immediately, whereas packets arriving just after a TTI has started, incur an additional delay of almost a TTI. Therefore, the jitter introduced by the physical layer is equal to $N_I T_R$. We call this the air interface queueing delay $T_{q,AI}$.

An MT accesses the UTRAN over the air interface via a base station (node B). Between node B and the ingress router R to the IP backbone network, the voice packets go through one or more radio network controllers (RNCs). Macro-diversity (soft hand-overs) requires a fast switching technology in the terrestrial radio access network. Therefore, the voice packets will traverse the UTRAN on top of AAL2 and the delay incurred in the UTRAN will be small (5-10 ms). We denote the minimum delay introduced in the UTRAN by $T_{m,UTRAN}$ and the queueing delay in the UTRAN by $T_{q,UTRAN}$.

3) *IP backbone network.*

Every voice packet has to queue in the IP backbone network. The distribution of the queueing delay in the backbone network depends on the scheduling strategy that is used. Assuming that voice packets receive HoL priority, the following accurate closed-form formula (derived in [6]) can be used for the (1-P)-quantile of the queueing delay,

$$T_{q,BB}(P) = A(H_{hl}, \rho, P) \frac{MTU_{voice,BB} + O_{PPP}}{R_{BB}} + H \frac{MTU_{data,BB} + O_{PPP}}{R_{BB}} \tag{6}$$

where

$$A(H_{hl},\rho,P)=\frac{\rho H_{hl}}{2(1-\rho)}+\left[Er^{-1}(H_{hl},P)-H_{hl}\right] \times\left[\left(\frac{\rho}{2(1-\rho)}\right)^2+\frac{\rho}{3(1-\rho)}\right]^{1/2} \tag{7}$$

In these expressions, ρ stands for the maximum voice load in the backbone network, H for the number of nodes a voice packet passes through from the ingress router to the VoIP gateway, and H_{hl} for the number of these nodes that are heavily loaded, that is, where the voice load reaches ρ. R_{BB} stands for the backbone link capacity, and $MTU_{data,BB}$ and $MTU_{voice,BB}$ stand for the maximum transfer units for data and for voice traffic in the backbone network, respectively. Finally, $Er^{-1}(H,P)$ stands for the (1-P)-quantile of a random variable with an Erlang distribution of H stages.

Apart from the queueing delay, voice packets also incur a processing delay and a serialisation delay in every node they cross in the backbone, and a propagation delay. We call the sum of the processing and the propagation delays the minimum backbone network delay $T_{m,BB}$. The total serialisation delay is given by $H\times B_F/R_{BB}$. We call R_{BB}/H the service rate of the backbone network.

4) Dejittering and decoding.

After having traversed the IP backbone network, the voice packets arrive at the VoIP gateway. In the dejittering buffer of the VoIP gateway, the first incoming voice packet is delayed for a certain amount of time, and subsequent voice packets are read out of the dejittering buffer periodically, at time instants that are $N_W T_W$ apart. This compensates for the delay jitter in the voice packet flow.

The larger the dejittering delay T_{jit}, the bigger the chance that voice packets will make it to the VoIP gateway in time to be read out by the codec [3]. If we want to limit the packet loss incurred in the dejittering buffer to $P_{loss,jit}$, we need to choose T_{jit} at least equal to

$$T_{jit}=T_{q,MT}+N_I T_R+T_{q,UTRAN}+T_{q,BB}(P_{loss,jit})\ . \tag{8}$$

The codec itself introduces a codec-dependent decoding delay that is approximately given by $T_{dec}=T_W$.

The last contribution to the mouth-to-ear delay comes from the PSTN, which introduces a small delay (propagation and switching). We call this delay the minimum PSTN delay $T_{m,PSTN}$.

5) Mouth-to-ear delay.

Consider an imaginary empty voice packet that incurs no interleaving delay (i.e., the imaginary case of $N_I=0$, $N_W=0$). The difference between an actual packet and this imaginary packet is that actual packets will incur additional delays due to packetisation, interleaving and serialisation.

The packetisation delay is given by $N_W T_W$, the sum of all the delay contributions due to interleaving is given by $3N_I T_R$, and the sum of all serialisation delays is given by B_F/R_S, where R_S is the effective service rate, given by

$$R_S=\frac{1}{\frac{1}{R_T}+H\frac{1}{R_{BB}}}\ . \tag{9}$$

The mouth-to-ear delay T_{M2E} is then given by

$$\begin{aligned} T_{M2E} &= T_0+3N_I T_R+T_{pack}+\frac{N_W B_W}{R_S} \\ &= T_0+3N_I T_R+N_W T_W\left(1+\frac{R_{cod}}{R_S}\right) \end{aligned} \tag{10}$$

where T_0 is the delay incurred by the imaginary packet. This delay can be calculated as

$$T_0=T_q+T_m+\frac{O_{RTP/UDP/IP}+O_{PPP}}{R_S}, \tag{11}$$

where T_m is the total minimum delay, excluding the interleaving delay and the delay needed to compensate for the jitter introduced by the air interface,

$$\begin{aligned} T_m &= T_{enc}+(T_{m,AI})_{N_I=0}+T_{m,UTRAN}+T_{m,BB} \\ &\quad +(T_{jit})_{N_I=0}+T_{dec}+T_{m,PSTN} \end{aligned} \tag{12}$$

and T_q is the total queueing delay, exclusive of the queueing delay at the air interface. Because the queueing delay in the backbone network is a stochastic variable, we use the (1-P)-quantile $T_{q,BB}(P)$ of its distribution function in our calculation of T_q,

$$T_q=T_{q,MT}+(T_{q,AI})_{N_I=0}+T_{q,UTRAN}+T_{q,BB}(P_{est}) \tag{13}$$

In this expression, P_{est} is the probability that our estimate of the queueing delay of the first packet in the backbone network underestimates the actual queueing delay.

If the first packet takes longer to arrive than we estimated, the method we use to set the parameters governing the operation of the VoIP application starts from data that are too optimistic. Therefore, P_{est} can be seen as the probability that the actual voice quality of the call will be worse than the quality we targeted when we set the VoIP application parameters.

Table 2. Minimum delays, queueing delays and service rates of the different stages in the example we considered.

	minimum delay	queueing delay	service rate
mobile terminal	20 ms	10 ms	n.a.
UMTS air interface	TTI	TTI	128 kbit/s
UTRAN	10 ms	0 ms	n.a.
IP backbone network	10 ms	1 ms	15.5 Mbit/s
VoIP gateway	TTI + 31 ms	n.a.	n.a.
PSTN	2 ms	0 ms	n.a.

B. Calculation of the packet loss probability

Assuming that only the UMTS air interface and the dejittering mechanism contribute to the packet loss, the packet loss probability P_{loss} is given by

$$P_{loss} = 1-(1-P_{loss,AI})(1-P_{loss,jit}) . \qquad (14)$$

Because we want to limit the number of voice packets arriving too late in the dejittering buffer, $P_{loss,jit}$ will usually be very small (much smaller than $P_{loss,AI}$) and P_{loss} will be roughly equal to $P_{loss,AI}$.

We assume that a voice packet is lost as soon as its encapsulating PPP frame contains erroneous bits. To calculate the probability that this happens, we assume that the residual bit error process on the transport channel over the air is fully described by the residual bit error rate RBER, that is, bit errors on the transport channel are statistically independent (because of the interleaving). RBER depends on the speed of the mobile user (which determines the Doppler-frequency characterising the fading on the wireless channel), the mean SNR per bit (E_b/N_0), and the number of radio frames N_I in each transmission time interval. To calculate RBER, we use the model we derived in [14]. The probability of one or more erroneous bits in an encapsulating PPP frame then becomes

$$P_{loss,AI} = 1-(1-\mathrm{RBER})^{B_F} . \qquad (15)$$

V. Setting the UMTS Air Interface Parameters, the Voice Packet Size and the Dejittering Delay

As an example we consider a scenario in which the mobile party travels at a speed of 60 km/h. The MT uses a GSM-EFR codec, of which the parameters are B_W=224 bits, T_W=20 ms and T_{LA}=0 ms. We assume that all echo is eliminated, that is, $EL=\infty$. We assume that voice packets receive HoL priority in the PPP layer of the air interface, and that fragmentation-oriented PPP is used, with the maximum transfer unit for data traffic chosen in such a way that a voice packet never has to queue for more than 10 ms. We assume that a header compression (HC) scheme is used that reduces the (average) number of RTP/UDP/IP overhead bits to 16 (the ROCCO algorithm [11] achieves this performance). The total overhead, including the 56 bits of the PPP overhead, is then given by 72 bits.

We assume that a transport channel with a bit rate of 128 kbit/s is available between the MT and the UTRAN, and that the standardised rate 1/3 convolutional code is used over the UMTS air interface. We assume that the delay incurred in the UTRAN is 10 ms. This covers processing, queueing, transmission, propagation and the dejittering necessary to make soft combining (macro- diversity) possible [2].

The IP backbone network is assumed to be a high-speed network with 155 Mbit/s links. Voice packets are assumed to traverse 10 nodes in the IP backbone, 5 of which are heavily loaded with voice. Voice packets receive HoL priority, and the maximum transfer units for voice and data are 440 and 1500 bytes, respectively. For the calculation of the queueing delay of the first packet we set $P_{est}=10^{-5}$. Under these assumptions, formula (6) gives $T_{q,BB}$=1 ms. The minimum delay, covering propagation (5 μs/km) and processing, is assumed to be 10 ms.

Finally we assume that the PSTN network introduces a delay of 2 ms (covering switching and propagation).

The minimum delays, queueing delays and service rates of the various transmission stages encountered by the voice packet flow are listed in Table 2. From the values of the queueing delays we deduce that we have to introduce a dejittering delay of at least 11 ms (plus the duration of the TTI on the air interface) if we want to keep the packet loss in the dejittering buffer below 10^{-5}.

The constants T_0 and R_S in the expression (10) for the mouth-to-ear delay can also be calculated using the data in Table 2. The total effective service rate R_S is approximately 128 kbit/s. The mouth-to-ear delay T_0 of an imaginary empty voice packet that incurs no interleaving delay is approximately 84.5 ms, and covers a total minimum delay T_m of 73 ms, a total queueing delay T_q of 11 ms, and a small serialisation delay for the header of 0.5 ms.

A. Trajectories followed by the working point

We can draw curves in the (T_{M2E}, I_e)-plane that correspond to working points that receive the same rating. We call these curves equirating curves. These curves have been plotted in Fig. 4, using the colour code of Table 1. The equirating curves corresponding to ratings of 60, 70 and 80 have been highlighted.

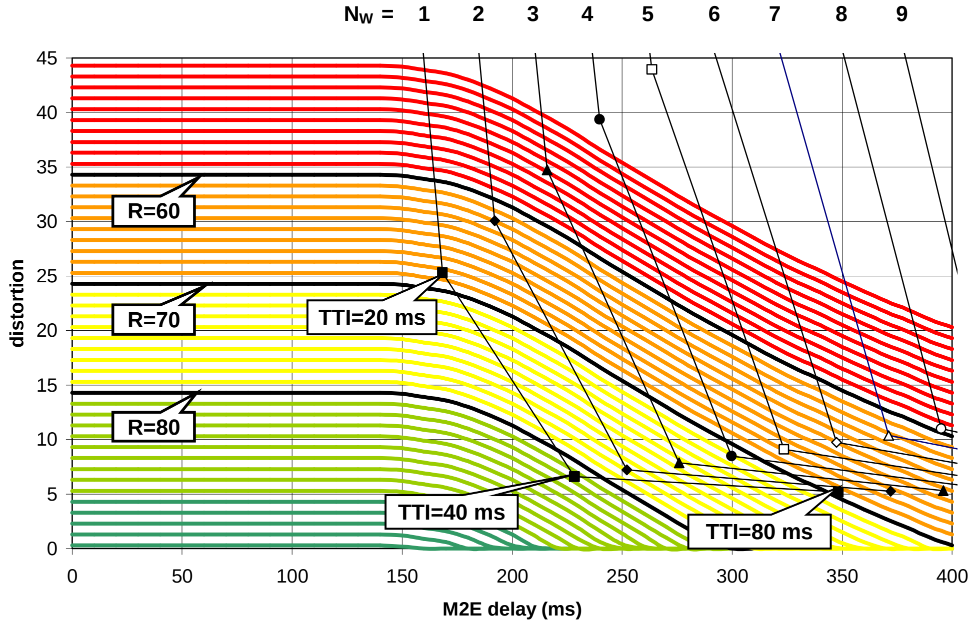

Fig. 4. Trajectories followed by the working point of the VoIP application. The mean area SNR per bit is 8 dB. We consider voice packets containing 1,2,...10 voice code words.

As we change the TTI from 10 ms to 20 ms, from 20 ms to 40 ms, and finally to 80 ms, the working point of the VoIP application follows a trajectory in the (T_{M2E}, I_e)-plane. We calculate several trajectories, one for each voice packet size we consider. If we assume the mean SNR per bit to be 8 dB, we get the trajectories depicted in Fig. 4. For a mean SNR per bit of 9 dB, we get the trajectories depicted in Fig. 5. The mean SNR per bit (E_b/N_0) has an important impact on the freedom we have in setting the VoIP application parameters and the UMTS air interface parameters. The region bounded by the equirating curve R=80 contains only one working point when (E_b/N_0) is 8 dB, and five when (E_b/N_0) is 9 dB.

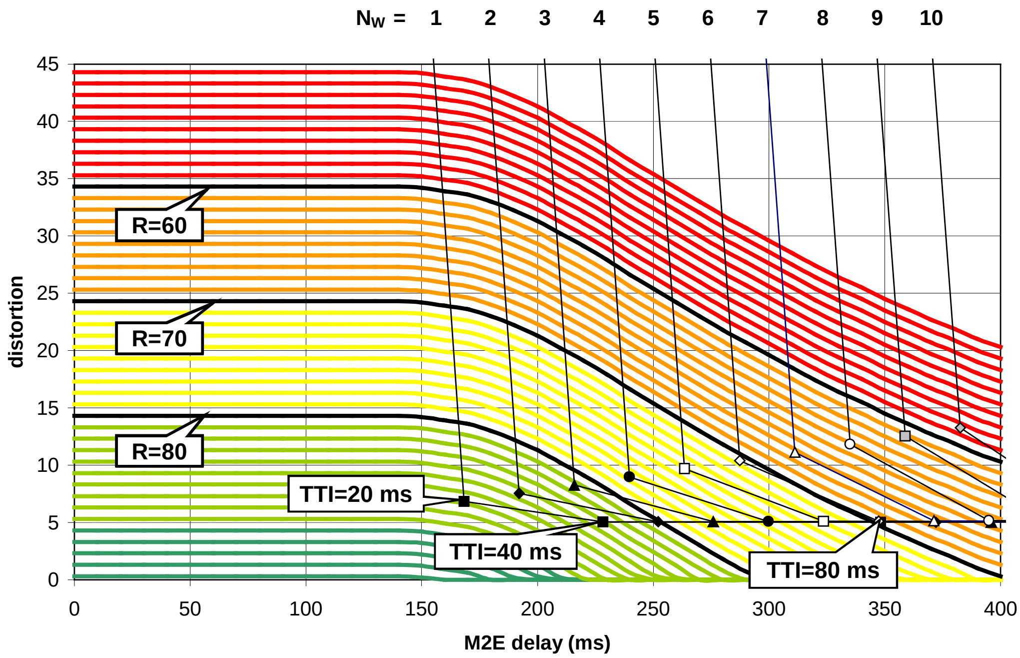

Fig. 5. Trajectories followed by the working point of the VoIP application. The mean area SNR per bit is 9 dB. We consider voice packets containing 1,2,...10 voice code words.

Table 3. Influence of the target voice quality and the per bit signal-to-noise ratio on the attainable IP layer bandwidth efficiency.

	target rating R=75			target rating R=80			target rating R=85		
E_b / N_0	N_W	filling factor		N_W	filling factor		N_W	Filling factor	
		HC	no HC		HC	no HC		HC	no HC
7.5 dB	1	77.2 %	39.3 %	-	-	-	-	-	-
8 dB	2	87.1 %	56.5 %	1	77.2 %	39.3 %	-	-	-
8.5 dB	3	91.0 %	66.1 %	1	77.2 %	39.3 %	-	-	-
9 dB	4	93.1 %	72.2 %	3	91.0 %	66.0 %	1	77.2 %	39.3 %
9.5 dB	5	94.4 %	76.4 %	4	93.1 %	72.2 %	2	87.1 %	56.5 %
10 dB	6	95.3 %	79.6 %	4	93.1 %	72.2 %	2	87.1 %	56.5 %

B. Determining the parameter values that yield the largest bandwidth efficiency for a given quality bound

The largest bandwidth efficiency is attained by packing as much voice words as possible into the same IP packet.

A typical value for the intrinsic quality of a call transported in G.711 format is 94. Assuming that we want the call to receive a rating of at least 80 (best or high quality), this means that the sum of the delay and the distortion impairments must remain smaller than 14.

The objective of attaining the highest possible bandwidth efficiency is then realised by selecting the rightmost trajectory in the (T_{M2E},I_e)-plane that passes through the region bounded by the equirating curve corresponding to R=80.

In Fig. 4, the second point, i.e. a TTI of 40 ms, on the first trajectory, i.e. N_W=1, qualifies. This means that a VoIP call that packs one code word in each voice packet and is served by a radio link that uses an interleaving span of 40 ms will receive a rating of more than 80. As Fig. 5 illustrates, for a power budget that is just 1 dB higher, the same point corresponds to a rating close to 85. We can now pack up to three code words into one voice packet and still achieve our target rating of 80.

We calculated the influence of the target voice quality, of the signal-to-noise ratio, and of header compression, on the filling factor that is attained over the wireless transport channel. The results are listed in Table 3. These results show that header compression increases the filling factor significantly. They also illustrate that a power budget increase can be used either to increase the bandwidth efficiency (keeping the same voice quality), or to increase the voice quality (keeping the same bandwidth efficiency).

VI. CONCLUSION

Guaranteeing end-to-end QoS to packetised voice calls requires respecting strict packet loss and delay bounds. We developed a method that allows us to determine the optimal values for the voice packet size and the UMTS air interface parameters simultaneously (the objective being the maximisation of the bandwidth efficiency).

Our study of the UMTS-to-PSTN scenario illustrates the value of the method. We found that with respect to bandwidth efficiency, the use of header compression is more important than the number of voice code words packed in each packet. However, the number of voice code words per packet still has an important impact on the bandwidth efficiency in the backbone network, where no header compression can be used. One could argue however that bandwidth efficiency for voice is not an issue in the backbone network, because the voice traffic volume is expected to be relatively small compared to the data traffic volume.

For a given codec type, we found that an increase of the power budget can be exploited either to increase the bandwidth efficiency, or to increase the quality of the voice call. The method we developed allows us to explore this trade-off.

REFERENCES

[1] M. Baker, “Speech Transport for Packet Telephony and Voice over IP”, Proc. of the SPIE Conference on Performance and Control of Network Systems III, Vol. 3842, Boston, September 1999.

[2] E. Dahlman, P. Benning, J. Knutsson, F. Ovesjö, M. Persson and C. Roobol, “WCDMA – The Radio Interface for Future Mobile Multimedia Communications”, IEEE Transactions on Vehicular Technology, Vol. 4, No. 4, pp. 1105-1118, November 1998.

[3] D. De Vleeschauwer, J. Janssen, G.H. Petit, “Delay Bounds for Low Bit Rate Voice Transport over IP Networks”, Proc. of the SPIE Conference on Performance and Control of Network Systems III, Vol. 3841, pp. 40-48, Boston, September 1999.

[4] D. De Vleeschauwer, J. Janssen, G.H. Petit, F. Poppe, “Quality Bounds for Packetized Voice Transport”, Alcatel Telecom Review, pp. 19-24, January 2000.

[5] D. De Vleeschauwer, J. Janssen, F. Poppe, G.H. Petit, “Choosing the Packet Size and Dejittering Delay for High-Quality Phone-to-Phone Calls Transported over an IP Backbone”, submitted to Performance Evaluation, November 1999.

[6] D. De Vleeschauwer, G.H. Petit, S. Wittewrongel, B. Steyaert and H. Bruneel, “Calculating the Dejittering Delay in Packet-Based Voice Networks with an Accurate Closed-Form Formula”, submitted to the European Journal of Operational Research, September 1999.

[7] ETSI Guide Draft 201 050, "Speech Processing, Transmission and Quality Aspects (STQ); Overall Transmission Plan Aspects for Telephony in a Private Network", November 1998.

[8] "The E-model, a computational model for use in transmission planning", ITU-T recommendation G. 107, May 2000.

[9] "Provisional Planning Values for the Equipment Impairment Factor I_e", Appendix to ITU-T Recommendation G.113 (Draft), September 1999.

[10] J. Janssen, R. Windey, D. De Vleeschauwer, G.H. Petit and J.-M. Leroy, "Maximum Delay Bounds for Voice Transported over Satellite Internet Access Networks", Proceedings of WOSBIS99, Rio De Janeiro, December 1999.

[11] L.-E. Jonsson, M. Degermark, H. Hannu, K. Svanbro, "RObust Checksum-based header COmpression scheme (ROCCO), draft-jonsson-robust-hc-02, October 1999.

[12] T. J. Kostas, M. S. Borella, I. Sidhu, G. M. Schuster, J. Grabiec, and J. Mahler, "Real-Time Voice Over Packet-Switched Networks", IEEE Network, Vol. 13, No. 1, pp 18-27, January/February 1998.

[13] J. Parantainen and S. Hamiti, "Delay Analysis for IP Speech over GPRS", Proc. of VTC 99, pp. 829-833, Amsterdam, September 1999.

[14] F. Poppe, D. De Vleeschauwer, G.H. Petit, "Guaranteeing Quality of Service to Packetised Voice over the UMTS Air Interface", Proceedings of the Eighth International Workshop on Quality of Service (IWQoS 2000), Pittsburgh, June 5-7, 2000.

[15] A.G. Valkó, A. Rácz and G. Fodor, "Voice QOS in Third Generation Mobile Systems", IEEE Journal on Selected Areas in Communications, Vol. 17, No. 1, pp. 109-123, January 1999.